

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2004, 10:43:16 ; Search time 92 Seconds

(without alignments)
10322.932 Million cell updates/sec

Title: US-09-662-454-3

Perfect score: 16009

Sequence: 1 MSTNPKPQRTKNTNRQP.....FPLCLMLLSVGVGIIYLLPNR 3010

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16009	100.0	3010	12	092969 hepatitis c
2	15949	99.6	3010	12	092970 hepatitis c
3	15937	99.6	3010	12	092971 hepatitis c
4	15909	99.4	3010	12	092828 hepatitis c
5	15887	99.2	3010	12	092972 hepatitis c
6	15772	98.5	3010	12	092829 hepatitis c
7	15392	96.1	3010	12	099AU2 hepatitis c
8	15381	96.1	3010	12	09QIX5 hepatitis c
9	15378	96.1	3010	12	09J3H7 hepatitis c
10	15358	95.9	3010	12	09J3H5 hepatitis c
11	15356	95.9	3010	12	09QIX6 hepatitis c
12	15338	95.8	3010	12	09J3H9 hepatitis c
13	15330	95.8	3010	12	P8996 hepatitis c
14	15330	95.8	3010	12	Q9DTE6 hepatitis c
15	15325	95.7	3010	12	Q9QIY4 hepatitis c
16	15323	95.7	3010	12	Q9J3I0 hepatitis c

17	15323	95.7	3010	12	09WXX2 hepatitis c
18	15322	95.7	3010	12	Q807P3 hepatitis c
19	15319	95.7	3010	12	Q9DTE9 hepatitis c
20	15303	95.6	3010	12	Q9DTE7 hepatitis c
21	15301	95.6	3010	12	Q9QIY3 hepatitis c
22	15299	95.6	3010	12	Q9J3I1 hepatitis c
23	15299	95.6	3010	12	Q9DTE4 hepatitis c
24	15280	95.4	3010	12	Q9J3H0 hepatitis c
25	15278	95.4	3010	12	Q9DTE9 hepatitis c
26	15273	95.4	3010	12	Q9DTE9 hepatitis c
27	15272	95.4	3010	12	Q9QIY6 hepatitis c
28	15267	95.4	3010	12	Q9J3G6 hepatitis c
29	15265	95.4	3010	12	Q81760 hepatitis c
30	15253	95.3	3010	12	Q8V638 hepatitis c
31	15253	95.3	3010	12	Q9J3V3 hepatitis c
32	15242	95.2	3010	12	Q9QIY5 hepatitis c
33	15239	95.2	3010	12	P90194 hepatitis c
34	15237	95.2	3010	12	Q9DTE7 hepatitis c
35	15237	95.2	3010	12	Q9DTE6 hepatitis c
36	15236	95.2	3010	12	Q9J3G1 hepatitis c
37	15235	95.2	3010	12	Q9DTE0 hepatitis c
38	15233	95.2	2864	12	Q9WIK8 hepatitis c
39	15225	95.1	3010	12	P90193 hepatitis c
40	15222	95.1	3010	12	Q9J3G8 hepatitis c
41	15221	95.1	3010	12	Q9QIX4 hepatitis c
42	15216	95.0	3010	12	Q9QIP06 hepatitis c
43	15215	95.0	3010	12	Q9J3F9 hepatitis c
44	15214	95.0	3010	12	Q68826 hepatitis c
45	15212.5	95.0	3013	12	Q9QIYC hepatitis c

ALIGNMENTS

RESULT 1

Q92969 PRELIMINARY; PRT; 3010 AA.
 AC Q92969;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HC-J4;
 RX MEDLINE=98240944; PubMed=9581788;
 RA Yaragi M., St Clare M., Shapiro M., Emerson S.U., Purcell R.H.,
 RA Bakh J.;
 RT "Transcripts of a chimeric clone of hepatitis C virus genotype 1b are
 RT infectious in vivo."
 RL Virology 244:161-172(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HC-J4;
 RA Yaragi M., Bakh J.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; AF054247; AAC15722.1; -;
 DR PIR; A61196; A61196.
 DR PIR; PQ0246; PQ0246.
 DR PIR; PS0329; PS0329.
 DR HSP; P26663; IUXP.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.

DR GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO:0003723; F:RNA binding; IEA.
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO:0005139; F:structural molecule activity; IEA.
 DR GO:0016740; F:transferase activity; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO:0006350; P:transcription; IEA.
 DR GO:0019079; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_cryptin.
 DR InterPro: IPR00410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR004109; Peptidase C29.
 DR InterPro: IPR007095; RNA_pol_DS_Ps.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00998; Viral RdRp; 1.
 DR P:Odor; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326776 MW; 9B3FD910CF00E2C5 CRC64;

Query Match 100.0%; Score 16009; DB 12; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTNPKPQKTKENTNRPPQDVKFPFGGQIVGGVYLLPRGPRLLGVRAATKASERSQPRG 60
 DB 1 MSTNPKPQKTKENTNRPPQDVKFPFGGQIVGGVYLLPRGPRLLGVRAATKASERSQPRG 60

QY 61 RROPIKARPEGRNAQPCYPWPLYGNEGLWAGWMLSPRGRSPSGPTDPRSRNLG 120
 DB 61 RROPIKARPEGRNAQPCYPWPLYGNEGLWAGWMLSPRGRSPSGPTDPRSRNLG 120

QY 121 KVIDTTCGPDALMGYIPLVGAFLGGAARALAHGVRLVDGVNATCNLPGCCFSIFLLA 180
 DB 121 KVIDTTCGPDALMGYIPLVGAFLGGAARALAHGVRLVDGVNATCNLPGCCFSIFLLA 180

QY 181 LLSCLTIPASAYEVRNVSGIHYHNTDCNSSIIVYEADVLIMHTPGVCPCVOEGNSSRCW 240
 DB 181 LLSCLTIPASAYEVRNVSGIHYHNTDCNSSIIVYEADVLIMHTPGVCPCVOEGNSSRCW 240

QY 241 ALTPTLAARNASVPTTIRRHVDLLVGTAAFCAMVVDGLCGSIFLVSQLFTFSPRRHET 300
 DB 241 ALTPTLAARNASVPTTIRRHVDLLVGTAAFCAMVVDGLCGSIFLVSQLFTFSPRRHET 300

QY 301 VQDCNSIYPGHVSHGRMAWMMNWSPTTALVVSQLLRIPQAVDMVACAHWGLVAGLA 360
 DB 301 VQDCNSIYPGHVSHGRMAWMMNWSPTTALVVSQLLRIPQAVDMVACAHWGLVAGLA 360

QY 361 YYSWGNWAKVILVALFAGVDGETHTTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420
 DB 361 YYSWGNWAKVILVALFAGVDGETHTTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420

QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGGCPRMASCRIPIWFAQGWGPIYTKPNSS 480

DB 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGGCPRMASCRIPIWFAQGWGPIYTKPNSS 480
 QY 481 DORPYCHYAPRCGVVPASQVGPVYCFPTSPVWVGTTDRSGVPTYSWGENETDVMNLN 540
 DB 481 DORPYCHYAPRCGVVPASQVGPVYCFPTSPVWVGTTDRSGVPTYSWGENETDVMNLN 540
 QY 541 NTRPPQGNWPGCTWNNSTGPTKTCGGPPCNIGVGNRTLICPTDCFRKHPAETVYTKCGSG 600
 DB 541 NTRPPQGNWPGCTWNNSTGPTKTCGGPPCNIGVGNRTLICPTDCFRKHPAETVYTKCGSG 600
 QY 601 PWLTPRCLVDYPVRLWHYPCTLNFSIFKRVYVGVGVEHRLNAACNWTGRONLEDRDRS 660
 DB 601 PWLTPRCLVDYPVRLWHYPCTLNFSIFKRVYVGVGVEHRLNAACNWTGRONLEDRDRS 660
 QY 661 ELSPLLLSTTEWQILPCAPTTLPALSTGLHLHQNIVDVQYLYGVGSFAFVFAIKWEYIL 720
 DB 661 ELSPLLLSTTEWQILPCAPTTLPALSTGLHLHQNIVDVQYLYGVGSFAFVFAIKWEYIL 720
 QY 721 LLELLILADARVCACLWMLLIAQAEAALENLVNNAASVAGAHGILSLFVFFCAAWYIKG 780
 DB 721 LLELLILADARVCACLWMLLIAQAEAALENLVNNAASVAGAHGILSLFVFFCAAWYIKG 780
 QY 781 RLAPGAAYAFYGVWPLLLILLALLPPRAYALDRMAASCGGAVLVGLVFLTLSPYKVFELT 840
 DB 781 RLAPGAAYAFYGVWPLLLILLALLPPRAYALDRMAASCGGAVLVGLVFLTLSPYKVFELT 840
 QY 841 RLILWLOYPITRAEAHMVQVWVPLNVRGORDAIILTCVAHPBELIFDITKLLAILGLPM 900
 DB 841 RLILWLOYPITRAEAHMVQVWVPLNVRGORDAIILTCVAHPBELIFDITKLLAILGLPM 900
 QY 901 VLQAGITRVYPYFVRAOGLTRACMLVRKVGAGHYVQVFMKLGALTCYYNNHLTPLRDWA 960
 DB 901 VLQAGITRVYPYFVRAOGLTRACMLVRKVGAGHYVQVFMKLGALTCYYNNHLTPLRDWA 960
 QY 961 HAGLRDLAVAVEPVVFSAMETKVIITWGAATAACGDIILGLPVARSARGKEIFLGPADSLEG 1020
 DB 961 HAGLRDLAVAVEPVVFSAMETKVIITWGAATAACGDIILGLPVARSARGKEIFLGPADSLEG 1020

QY 1021 QGRLLEAPITAYSQOQTRGVLCIITSLTGRDNQVEGEVQVWVSTATQSLAFCINGVCWT 1080
 DB 1021 QGRLLEAPITAYSQOQTRGVLCIITSLTGRDNQVEGEVQVWVSTATQSLAFCINGVCWT 1080

QY 1081 VHGAGSKTLAGKGPITOMYINVDLILVGOAPPGARSMTPCSCSSDLYLVTRHADVI 1140
 DB 1081 VHGAGSKTLAGKGPITOMYINVDLILVGOAPPGARSMTPCSCSSDLYLVTRHADVI 1140

QY 1141 PVRRRGDSRGSLLSPRPVSYLKGSGGPLLCPSGHVGVFRAAVCTRGVAKAVDFTPVES 1200
 DB 1141 PVRRRGDSRGSLLSPRPVSYLKGSGGPLLCPSGHVGVFRAAVCTRGVAKAVDFTPVES 1200

QY 1201 METTVRSPTVNTSTPPAVPQTFQVAHHAFTSGSKSTKVPAAAYAAQGYKVLVLPNSVAA 1260
 DB 1201 METTVRSPTVNTSTPPAVPQTFQVAHHAFTSGSKSTKVPAAAYAAQGYKVLVLPNSVAA 1260

QY 1261 TLGFGAYMSKAGIDPNRTGVRTITTCGSIITYSYGKFLADGGCGGAYDIIICDECHS 1320
 DB 1261 TLGFGAYMSKAGIDPNRTGVRTITTCGSIITYSYGKFLADGGCGGAYDIIICDECHS 1320

QY 1321 TDSSTILGIGTVDQAEATAGARLVLAATATPPGSVTVPHNIEELGLSNNGEIPYGKAI 1380
 DB 1321 TDSSTILGIGTVDQAEATAGARLVLAATATPPGSVTVPHNIEELGLSNNGEIPYGKAI 1380

QY 1381 PIEATKGRHILFCHSKKCDLAAKLTGLGLNNAVYTRGLDVSVPPIGDVVVATDAL 1440
 DB 1381 PIEATKGRHILFCHSKKCDLAAKLTGLGLNNAVYTRGLDVSVPPIGDVVVATDAL 1440

QY 1441 MTGFTGDPDSVLDNCTVQTVDFSLDPTFTTETTTVPQDAVRSQRGRGRCGSGIYR 1500
 DB 1441 MTGFTGDPDSVLDNCTVQTVDFSLDPTFTTETTTVPQDAVRSQRGRGRCGSGIYR 1500

QY 1501 FYTPGSRPGRFSDSSVLCECAGAWYELTPAETSVELRAYLNTPLPVQCQDHLFEWES 1560

Db 1501 FVTPGERSGMDSSVLCCEVDAGCANVELTPTAETSVRLRAYLNTPGLPVQDHLFPWES 1560
Qy 1561 VFTGLTHIDAHFLSQTQAGDNFPYLVAAYQATVCARAQAAPPSPDQWKKLIRLKTPLHG 1620
Db 1561 VFTGLTHIDAHFLSQTQAGDNFPYLVAAYQATVCARAQAAPPSPDQWKKLIRLKTPLHG 1620
Qy 1621 PTELLYRIGAVQNEVILTHPTIKYIMACWSADLEVVTTWLVGVZLAALAAAYCLTTGSV 1680
Db 1621 PTELLYRIGAVQNEVILTHPTIKYIMACWSADLEVVTTWLVGVZLAALAAAYCLTTGSV 1680
Qy 1691 VIVGRIILSGPAPVVPDEVLVYQSEDEMECASQLPYTEOGWQIAEOFKQKALGLLOTAT 1740
Db 1691 VIVGRIILSGPAPVVPDEVLVYQSEDEMECASQLPYTEOGWQIAEOFKQKALGLLOTAT 1740
Qy 1741 KQAEAAAPVVEKRALETFWAKHMWFIISGIVLAGLSTIPGNPAIASLMAFTASITSP 1800
Db 1741 KQAEAAAPVVEKRALETFWAKHMWFIISGIVLAGLSTIPGNPAIASLMAFTASITSP 1800
Qy 1801 LTTQNTLILFNILGWNAAQLAPPASAFVAGIAGAAVSGIGLKVLDILAGYGAV 1860
Db 1801 LTTQNTLILFNILGWNAAQLAPPASAFVAGIAGAAVSGIGLKVLDILAGYGAV 1860
Qy 1861 GALVAFKVMGSEVESTDLNLLPAILSPGALVGVVCAAILRRHVGPGEAGVOMNRLLI 1920
Db 1861 GALVAFKVMGSEVESTDLNLLPAILSPGALVGVVCAAILRRHVGPGEAGVOMNRLLI 1920
Qy 1921 AFASRGNHVSPTHVPSDAAARVTOQLSSITITQLLKLHQWINECDSTPCSGSLRDV 1980
Db 1921 AFASRGNHVSPTHVPSDAAARVTOQLSSITITQLLKLHQWINECDSTPCSGSLRDV 1980
Qy 1981 WDMICTVLTDPKWLQSKLLPRLGVPFLSCQRYKGWVRGDGIMQTCPCGAQIAGHV 2040
Db 1981 WDMICTVLTDPKWLQSKLLPRLGVPFLSCQRYKGWVRGDGIMQTCPCGAQIAGHV 2040
Qy 2041 NGSMRIIVGPRTCSTWGTGTPINATYTCPTSPAPNYSRALMVAABEYVEVTRVGDH 2100
Db 2041 NGSMRIIVGPRTCSTWGTGTPINATYTCPTSPAPNYSRALMVAABEYVEVTRVGDH 2100
Qy 2101 YVTGMTDNDKVCQVPAPEFFTEVDGVLHRYAPACKPLLEDVTFQVGLNQVLVGSQ 2160
Db 2101 YVTGMTDNDKVCQVPAPEFFTEVDGVLHRYAPACKPLLEDVTFQVGLNQVLVGSQ 2160
Qy 2161 PCPEPDVTVLTSMLTDPSTHATAETARRIARGSPRLASSASQLSAPSLKATCTTHHD 2220
Db 2161 PCPEPDVTVLTSMLTDPSTHATAETARRIARGSPRLASSASQLSAPSLKATCTTHHD 2220
Qy 2221 SPADLIEANLLWRQMGCNITRVESENKVVILDSFPLHAEGDERISVAABELKRSK 2280
Db 2221 SPADLIEANLLWRQMGCNITRVESENKVVILDSFPLHAEGDERISVAABELKRSK 2280
Qy 2281 FPSALPIWARPDYNPDLLESWKDDPYVPVHGCPLPPTKAPLIPPPRRKRTVVLTESV 2340
Db 2281 FPSALPIWARPDYNPDLLESWKDDPYVPVHGCPLPPTKAPLIPPPRRKRTVVLTESV 2340
Qy 2341 SSALAEIATKTFSSGSSAVDGTATLPDLASDDGKGSDESYSMPLEGPDPDL 2400
Db 2341 SSALAEIATKTFSSGSSAVDGTATLPDLASDDGKGSDESYSMPLEGPDPDL 2400
Qy 2401 SDGWSVTSEASEDVVCCSMSTWTGALITPCAABESKLPINPLNSLIRHNNVYAT 2460
Db 2401 SDGWSVTSEASEDVVCCSMSTWTGALITPCAABESKLPINPLNSLIRHNNVYAT 2460
Qy 2461 SRASLRQKVTDRQLQVLDHHDYADV.KEMKAKASTVAKLLSIEACKLTPPHSAKSF 2520
Db 2461 SRASLRQKVTDRQLQVLDHHDYADV.KEMKAKASTVAKLLSIEACKLTPPHSAKSF 2520
Qy 2521 GYKAKDVRLSSRAVNHRSVWEDLLEDTETPTDITIMAKSEVFCVQPEKGRKPARLIV 2580
Db 2521 GYKAKDVRLSSRAVNHRSVWEDLLEDTETPTDITIMAKSEVFCVQPEKGRKPARLIV 2580
Qy 2581 FPDGLGVRCERKALYDVVSTLPAVWSSYGFQYSPKQRFVFLVNTWKSCKPMGFSYDT 2640
Db 2581 FPDGLGVRCERKALYDVVSTLPAVWSSYGFQYSPKQRFVFLVNTWKSCKPMGFSYDT 2640

Qy 2641 RCFDSTVTESDIRVEESIYOCDDIAPARQAIRSLTBRLYIGGFLTNKGNCGYRRCRA 2700
Db 2641 RCFDSTVTESDIRVEESIYOCDDIAPARQAIRSLTBRLYIGGFLTNKGNCGYRRCRA 2700
Qy 2701 SGVLTTSCTGNTLTCYLKATAACRAAKLQDCIMLVNGDDLVVICESAGTQEDAAALRAFTE 2760
Db 2701 SGVLTTSCTGNTLTCYLKATAACRAAKLQDCIMLVNGDDLVVICESAGTQEDAAALRAFTE 2760
Qy 2761 AMTRYSAAPPDGPDPPEVDLELITSCSSNSVVAHDASGRVYVYTRDPTPLARAAMETAR 2820
Db 2761 AMTRYSAAPPDGPDPPEVDLELITSCSSNSVVAHDASGRVYVYTRDPTPLARAAMETAR 2820
Qy 2821 HTPINSMLGNIIMYAPTLWARMILMTHFFSILLAQEOLKALDCQIYGACYSIEPLDLPQ 2880
Db 2821 HTPINSMLGNIIMYAPTLWARMILMTHFFSILLAQEOLKALDCQIYGACYSIEPLDLPQ 2880
Qy 2881 IIERLHGLSAFTLHYSFGPEINRVASCLRKLGVPPLRTWHRARSVRKLLSQGGRATC 2940
Db 2881 IIERLHGLSAFTLHYSFGPEINRVASCLRKLGVPPLRTWHRARSVRKLLSQGGRATC 2940
Qy 2941 GRVLFNVAWVTKLTPIPAASQDLISGWTVAAGYSGGDIYHSLSRAPRPFPICLLLSV 3000
Db 2941 GRVLFNVAWVTKLTPIPAASQDLISGWTVAAGYSGGDIYHSLSRAPRPFPICLLLSV 3000
Qy 3001 GVGIYLLPNR 3010
Db 3001 GVGIYLLPNR 3010
RESULT 2
O92970 PRELIMINARY; PRT; 3010 AA.
AC O92970;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J4;
RX MEDLINE=98240944; PubMed=9581788;
RA Yanagi M., St. Clare M., Shapiro M., Emerson S.U., Purcell R.H.,
RT Bakh J.;
RT "Transcripts of a chimeric clone of hepatitis C virus genotype 1b are
infectious in vivo."
RL Virology 244:161-172(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J4;
RA Yanagi M., Bakh J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DDJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF054248; AAC15723.1; -.
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PS0329; PS0329.
DR HSSP; P26663; 1JXP.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:vital envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

GO: GO:0005198; F:Structural molecule activity; IEA.
GO: GO:0016740; F:transferase activity; IEA.
GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
GO: GO:0006350; P:transcription; IEA.
GO: GO:0019079; P:virial genome replication; IEA.
GO: GO:0019079; P:virial genome replication; IEA.
DR InterPro: IPR009003; Cys Ser trypsin.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR000745; HCV NS4a.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRp.
DR InterPro: IPR004109; Peptidase C29.
DR InterPro: IPR007095; RNA pol DS PS.
DR InterPro: IPR007094; RNA pol PSvir.
DR Pfam: PF01543; HCV capsid; 1.
DR Pfam: PF01542; HCV core; 1.
DR Pfam: PF01539; HCV env; 1.
DR Pfam: PF01560; HCV NS1; 1.
DR Pfam: PF01538; HCV NS2; 1.
DR Pfam: PF02907; HCV NS3; 1.
DR Pfam: PF01006; HCV NS4a; 1.
DR Pfam: PF01001; HCV NS4b; 1.
DR Pfam: PF01506; HCV NS5a; 1.
DR Pfam: PF00998; Viral RdRp; 1.
DR ProDom: PD186662; HCV NS1; 1.
DR SMART; SM00487; DEXDc; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 326840 MW; 45FB399AD6141A88 CRC64;

Query Match 99.6%; Score 15949; DB 12; Length 3010;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2998; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
1 MSTNPKPQKTRNTNRNRRPDQYKFGGGQIVGGVLLPRGRLGVRAATRKASERSQPRG 60
1 MSTNPKPQKTRNTNRNRRPDQYKFGGGQIVGGVLLPRGRLGVRAATRKASERSQPRG 60
61 RQPIKARRPGRWAQAGYIPWLYGNEGLWAGWLLSPGRSRPSWGPDPRRSRNLG 120
61 RQPIKARRPGRWAQAGYIPWLYGNEGLWAGWLLSPGRSRPSWGPDPRRSRNLG 120
121 KVIDLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFIFLLA 180
121 KVIDLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFIFLLA 180
181 LLSCLTIPASAYEVNRVNSGIYHVNDNCNSSITVEAADVIMHTPCVPCVQSGNSRCWV 240
181 LLSCLTIPASAYEVNRVNSGIYHVNDNCNSSITVEAADVIMHTPCVPCVQSGNSRCWV 240
241 ALTPPLAARNASVPTTIRRHVDLLVGTAAFCASMYVGDLCGSIPLVSLFTFSRPHET 300
241 ALTPPLAARNASVPTTIRRHVDLLVGTAAFCASMYVGDLCGSIPLVSLFTFSRPHET 300
301 VODKNCSTYIPGVSHGRWADMMNWSPTTALVWSQLRIPQAVVDMVAGAHWGLAGLA 360
301 VODKNCSTYIPGVSHGRWADMMNWSPTTALVWSQLRIPQAVVDMVAGAHWGLAGLA 360
361 YYSVMGNWAKVLI VALLFAGVDGETHITGRVAGHTTSGFTSLFSSGASQKIQLVNTNGSW 420
361 YYSVMGNWAKVLI VALLFAGVDGETHITGRVAGHTTSGFTSLFSSGASQKIQLVNTNGSW 420
421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERWASCRPDWFAQGWGPTIYTKPNSS 480
421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERWASCRPDWFAQGWGPTIYTKPNSS 480
481 DQRPYCHWYAPRCGVVWVASFVCGGVYCFPTSPVVVGTDRSGVFTYSWGENETDMLLN 540

481 DQRPYCHWYAPRCGVVWVASFVCGGVYCFPTSPVVVGTDRSGVFTYSWGENETDMLLN 540
541 NTRPQGNWFGCTWMNSTGFTKTCGGPCNIGGVGNRTLIPTDCFRKHPATYTKCSG 600
541 NTRPQGNWFGCTWMNSTGFTKTCGGPCNIGGVGNRTLIPTDCFRKHPATYTKCSG 600
601 PWLTPRCCLVDYPRYLWHYPCILNPSIFKVRMYGVGVEHRLNAA CNMTGERCNLEDRRS 660
601 PWLTPRCCLVDYPRYLWHYPCILNPSIFKVRMYGVGVEHRLNAA CNMTGERCNLEDRRS 660
661 ELSPLLLSTTEWQILPCAFTTLPALSTGLIHLHQNIVDVQVLYGVGSFAVFAIKWEYL 720
661 ELSPLLLSTTEWQILPCAFTTLPALSTGLIHLHQNIVDVQVLYGVGSFAVFAIKWEYL 720
721 LFLLLADARVCACILMMMLLIAQAEEALENLVNLNAA SVAGAHGILSLVFPFCAWYIKG 780
721 LFLLLADARVCACILMMMLLIAQAEEALENLVNLNAA SVAGAHGILSLVFPFCAWYIKG 780
781 RLAPGAAYAFYGVWPLLLLLALPPRAVALDREMAASCGGAVLVGLVFLTISPYKVFILT 840
781 RLAPGAAYAFYGVWPLLLLLALPPRAVALDREMAASCGGAVLVGLVFLTISPYKVFILT 840
841 ELIWMLOVFIITRAEAMQVWVPLNVRGGRDAIILLTCAVHPELIFDITKLLAILGFLM 900
841 ELIWMLOVFIITRAEAMQVWVPLNVRGGRDAIILLTCAVHPELIFDITKLLAILGFLM 900
901 VLOAGITRVPYFVRAOGLIRACMLVRKAGHYVQMFKLGALTGTIVYVNHLPDRWA 960
901 VLOAGITRVPYFVRAOGLIRACMLVRKAGHYVQMFKLGALTGTIVYVNHLPDRWA 960
961 HAGRLDVAVPEVVPVFSAMETKVITWGAADTAACGDITGLPVSARRGKEIFLGADSLEG 1020
961 HAGRLDVAVPEVVPVFSAMETKVITWGAADTAACGDITGLPVSARRGKEIFLGADSLEG 1020
1021 QGWLLAPITAYSQOTRGVLCIIITSLTRDKNQEVEGVVSTATQSLATCINGVCWT 1080
1021 QGWLLAPITAYSQOTRGVLCIIITSLTRDKNQEVEGVVSTATQSLATCINGVCWT 1080
1081 VYHGAGSKTLAGKPGPIQMTYTNVDLLVHQAPPGARSMTPCSGSSDLYLVTRHADVI 1140
1081 VYHGAGSKTLAGKPGPIQMTYTNVDLLVHQAPPGARSMTPCSGSSDLYLVTRHADVI 1140
1141 PVRRGSRGSLSPRPVSYLKSGGPPLLCPSGHVGVFRAA VCTRGVAKAVDFIPVES 1200
1141 PVRRGSRGSLSPRPVSYLKSGGPPLLCPSGHVGVFRAA VCTRGVAKAVDFIPVES 1200
1201 METTWRSVFTDNSTPPAVPOTFQVAHLHAPTSGKSTKVPAA YAAQYKVLNPSVAA 1260
1201 METTWRSVFTDNSTPPAVPOTFQVAHLHAPTSGKSTKVPAA YAAQYKVLNPSVAA 1260
1261 TLGFGAYMSKAGHDPIRTGVRTITTCGSIITYGKFLADGGSGGADYIIICDECHS 1320
1261 TLGFGAYMSKAGHDPIRTGVRTITTCGSIITYGKFLADGGSGGADYIIICDECHS 1320
1321 TDSITLIGITVLDQAEATAGARLVLTATATPPGSVTVPHPNIEIGLSNNGEIPFYGKAI 1380
1321 TDSITLIGITVLDQAEATAGARLVLTATATPPGSVTVPHPNIEIGLSNNGEIPFYGKAI 1380
1381 PIEAIKGRHLIFCHSKKKCDLAAKLTGLGLNNAVAYRGLDVSVIPPIGVVVVATDAL 1440
1381 PIEAIKGRHLIFCHSKKKCDLAAKLTGLGLNNAVAYRGLDVSVIPPIGVVVVATDAL 1440
1441 MTGTFGPDSDVIDCNTCVTVDFSLDPTTIEFTTTPQDVSRSQRGRGTGRSGSIYR 1500
1441 MTGTFGPDSDVIDCNTCVTVDFSLDPTTIEFTTTPQDVSRSQRGRGTGRSGSIYR 1500
1501 FVTPGERPSGFMFSSVLCCEYDAGCAWVELTPASTSVLRAYLNTPGLPVQCDELFWES 1560
1501 FVTPGERPSGFMFSSVLCCEYDAGCAWVELTPASTSVLRAYLNTPGLPVQCDELFWES 1560
1561 VFTGJTHIDAHFISQTKOAGDNFFYLVAQATVCARAAQAPPPSWDQWKKLIRKPLTHG 1620

1561 VFTGLTHDAHFLSQTQAGNPFYLVAYQATVCARAQAPPSSWDQWKCLIRLKPETHG 1620
 1621 PTELLVRLGAVQNEVILTHPTIKYIMACMSADLEVTTSTWLVGGVLAALAAACLTGSSV 1680
 1621 PTELLVRLGAVQNEVILTHPTIKYIMACMSADLEVTTSTWLVGGVLAALAAACLTGSSV 1680
 1681 VIVGRILISGKPAVVPDREVLYOEFDEMERCASQLPVIEQGMQLAQFQKAGLGLQTAT 1740
 1681 VIVGRILISGKPAVVPDREVLYOEFDEMERCASQLPVIEQGMQLAQFQKAGLGLQTAT 1740
 1741 KQBAAPVSVESKWRALETWAKHMNFISGIQYLAGSLTLPNGPAIASLMTASITSP 1800
 1741 KQBAAPVSVESKWRALETWAKHMNFISGIQYLAGSLTLPNGPAIASLMTASITSP 1800
 1801 LTTQNTLLFNILGQWAAQALPASAFAVAGIAGAAVSGIGLQKVLVDILAGYGAGVA 1860
 1801 LTTQNTLLFNILGQWAAQALPASAFAVAGIAGAAVSGIGLQKVLVDILAGYGAGVA 1860
 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEVQWNRLLI 1920
 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEVQWNRLLI 1920
 1921 AFASRGNHVSPTHYVPESDAAARVQILSSLTITOLLKRLHOMINEDCSTPCSGSLRDV 1980
 1921 AFASRGNHVSPTHYVPESDAAARVQILSSLTITOLLKRLHOMINEDCSTPCSGSLRDV 1980
 1981 WDWICTVLTDFTKWLQSKLPLRPGVFPFLSCQRYGKGVWGDGIMQTCPCGACIAGHVK 2040
 1981 WDWICTVLTDFTKWLQSKLPLRPGVFPFLSCQRYGKGVWGDGIMQTCPCGACIAGHVK 2040
 2041 NGSMRIVGPTCSNTWHGTPIINAYTTGCTPSPAPNYSRALMRVAEAEVVEVTRVGDHF 2100
 2041 NGSMRIVGPTCSNTWHGTPIINAYTTGCTPSPAPNYSRALMRVAEAEVVEVTRVGDHF 2100
 2101 YVTGMTIDNKKCCQVPAPPEFFTEVDGVRHLRVAPACKPLLRDVTQVGLNQYLVGSQ 2160
 2101 YVTGMTIDNKKCCQVPAPPEFFTEVDGVRHLRVAPACKPLLRDVTQVGLNQYLVGSQ 2160
 2161 PCPEPDVTVLTMIDPSSHITAKRRLARGSPPSLASSASQISAKCTCTTHHD 2220
 2161 PCPEPDVTVLTMIDPSSHITAKRRLARGSPPSLASSASQISAKCTCTTHHD 2220
 2221 SPDADLIEANLLMRQEMGNITRVESENKVVILDSFEPLHAEGDEREISVAAELTRKSRK 2280
 2221 SPDADLIEANLLMRQEMGNITRVESENKVVILDSFEPLHAEGDEREISVAAELTRKSRK 2280
 2281 FPSALPIWARPDPNPPLESWKDPDYVPVVGCPLEPTTKAPFIPPPRRKRTVVLTSNV 2340
 2281 FPSALPIWARPDPNPPLESWKDPDYVPVVGCPLEPTTKAPFIPPPRRKRTVVLTSNV 2340
 2341 SSALAEALATKTFSSGSSAVDSGTATLPDLASDDGKGDVSESYSNMPLEGEQPDIL 2400
 2341 SSALAEALATKTFSSGSSAVDSGTATLPDLASDDGKGDVSESYSNMPLEGEQPDIL 2400
 2401 SDGSWSTVSEASEDVVCCSMYSWTGALITPCAAEESKLPINPLSNLLRHNNMVYATT 2460
 2401 SDGSWSTVSEASEDVVCCSMYSWTGALITPCAAEESKLPINPLSNLLRHNNMVYATT 2460
 2461 SRASLRQKVTDRLOVLDHVDVLEKMKAKASTYKAKLSTEEACKLTPPHSAKSF 2520
 2461 SRASLRQKVTDRLOVLDHVDVLEKMKAKASTYKAKLSTEEACKLTPPHSAKSF 2520
 2521 GYGAKDVNLSRRVNHRSVWEDLLDTEPTIDTTIMAKSEVPCVQKGGKPARLIV 2580
 2521 GYGAKDVNLSRRVNHRSVWEDLLDTEPTIDTTIMAKSEVPCVQKGGKPARLIV 2580
 2581 FPDGLGVRCCKMALDYVVTSLPQAVMGSSYGFQYSPKQRFVFLVNTWKSCKCPMGFSYDT 2640
 2581 FPDGLGVRCCKMALDYVVTSLPQAVMGSSYGFQYSPKQRFVFLVNTWKSCKCPMGFSYDT 2640
 2641 RCFDSTVTESDIRVEESYCCDLAPARQAIRSLTERLYIGGLTNSKQNCGRYCRRA 2700
 2641 RCFDSTVTESDIRVEESYCCDLAPARQAIRSLTERLYIGGLTNSKQNCGRYCRRA 2700

2701 SGVLTSCGNTLTCYLKATAACRAAKIQDCTMLVNGDDLVVICSAGTQEDAAALRAFTE 2760
 2701 SGVLTSCGNTLTCYLKATAACRAAKIQDCTMLVNGDDLVVICSAGTQEDAAALRAFTE 2760
 2761 AMTRYSAPGDPDPQPEYDLELITSCSSNVSVADHASKREVVYLTTRDPTTFLARAAMETAR 2820
 2761 AMTRYSAPGDPDPQPEYDLELITSCSSNVSVADHASKREVVYLTTRDPTTFLARAAMETAR 2820
 2821 HTPINSWLNIIIMYAPTLWARMLMTHFFSILLAQBLEKALDCQIYGACYSIEPLDLQ 2880
 2821 HTPINSWLNIIIMYAPTLWARMLMTHFFSILLAQBLEKALDCQIYGACYSIEPLDLQ 2880
 2881 IIERLHGLSAFTLHVSYPGEINRVASCLRKIGVPPPLRTWRHARSVRAKLLSQGGRAATC 2940
 2881 IIERLHGLSAFTLHVSYPGEINRVASCLRKIGVPPPLRTWRHARSVRAKLLSQGGRAATC 2940
 2941 GRYLFNVAVRTKLKUTPIPAASQDLSGWFVAGYSGGDIYHLSLRARPEWFFCLILLSV 3000
 2941 GRYLFNVAVRTKLKUTPIPAASQDLSGWFVAGYSGGDIYHLSLRARPEWFFCLILLSV 3000
 3001 GYGIYLLPNR 3010
 3001 GYGIYLLPNR 3010

RESULT 3
 092971 PRELIMINARY; PRT; 3010 AA.

AC 092971;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 CX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HC-J4;
 RX MEDLINE=98240944; PubMed=9581788;
 RA Yanagi M., St Clare M., Shagiro M., Emerson S.U., Purcell R.H.,
 RA Bukh J.;
 RT "Transcripts of a chimeric clone of hepatitis C virus genotype 1b are
 RT infectious in vivo."
 RL Virology 244:161-172(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HC-J4;
 RA Yanagi M., Bukh J.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; AF054249; AAC15724.1; -;
 DR PIR; A61196; A61196.
 DR PIR; P02046; P02046.
 DR PIR; P08004; P08004.
 DR PIR; P03029; P03029.
 DR HSSP; P26663; 1JXP.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.

DR	GO:	GO:0006350;	P:transcription; IEA.
DR	GO:	GO:0019079;	P:viral genome replication; IEA.
DR	GO:	GO:0019087;	P:viral transformation; IEA.
DR	InterPro:	IPR009003;	Cys_Ser_trypsin.
DR	InterPro:	IPR001410;	DEAD.
DR	InterPro:	IPR002522;	HCV_capsid.
DR	InterPro:	IPR002521;	HCV_core.
DR	InterPro:	IPR002519;	HCV_env.
DR	InterPro:	IPR002531;	HCV_NS1.
DR	InterPro:	IPR002515;	HCV_NS2.
DR	InterPro:	IPR000745;	HCV_NS4a.
DR	InterPro:	IPR001490;	HCV_NS4b.
DR	InterPro:	IPR002868;	HCV_NS5a.
DR	InterPro:	IPR002166;	HCV_RdRp.
DR	InterPro:	IPR004109;	Peptidase_C29.
DR	InterPro:	IPR007095;	RNA_pol_DS_PS.
DR	InterPro:	IPR007094;	RNA_pol_PSVir.
DR	Fam:	PF01543;	HCV_capsid; 1.
DR	Fam:	PF01542;	HCV_core; 1.
DR	Fam:	PF01539;	HCV_env; 1.
DR	Fam:	PF01560;	HCV_NS1; 1.
DR	Fam:	PF01538;	HCV_NS2; 1.
DR	Fam:	PF02907;	HCV_NS3; 1.
DR	Fam:	PF01066;	HCV_NS4a; 1.
DR	Fam:	PF01001;	HCV_NS4b; 1.
DR	Fam:	PF01506;	HCV_NS5a; 1.
DR	Fam:	PF00998;	Viral_RdRp; 1.
DR	ProDom:	PD186062;	HCV_NS1; 1.
DR	SMART:	SM00487;	DEXdc; 1.
DR	KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;	
DR	KW	Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.	
DR	SQ	SEQUENCE	3010 AA; 326833 MW; 8FBE0CFB99CBF82 CRC64;
<hr/>			
Query Match 99.6%; Score 15937; DB 12; Length 3010;			
Best Local Similarity 99.4%; Pred. No. 0;			
Matches 2993; Conservative 9; Mismatches 8; Indels 0; Gaps 0;			
<hr/>			
QY	1	MSTNPKPQRKT	KNTNNRRPDVXPPGGGQIVGGVYLPRRGPLGVRAATRKASERSQPRG 60
DB	1	MSINPKPQRKT	KNTNNRPDVAFPGGGQIVGGVYLFRGPRLGVRATKTSERSQPRG 60
<hr/>			
QY	61	RRIPIPKAREPEGRAWQ	QGYPNPLYNGNEGLGWAGLLSPRGSPSPSGPTDPRRSRLG 120
DB	61	RRIPIPKAREPEGRAWQ	QGYPNPLYNGNEGLGWAGLLSPRGSPSPSGPTDPRRSRLG 120
<hr/>			
QY	121	KVIDTLTCGFADLMGYI	PLVCAPLGGGAARALAHGVVEGVNATGNLPCCSFSIFLLA 180
DB	121	KVIDTLTCGFADLMGYI	PLVCAPLGGGAARALAHGVRLDEGVNATGNLPCCSFSIFLLA 180
<hr/>			
QY	181	LLSCLTTPASAYEVRNV	SGIYHVHTNDSCNSIVVEAADVMHTEPCGVPVCEGNSSRCWV 240
DB	181	LLSCLTTPASAYEVRNV	SGIYHVHTNDSCNSIVVEAADVMHTEPCGVPVCEGNSSRCWV 240
<hr/>			
QY	241	ALPTPLAARNASVPTTI	IRRHVDLLVGTAAPCSAMVGDLCGSIFLVSQLFTPSPRHET 300
DB	241	ALPTPLAARNASVPTTI	IRRHVDLLVGTAAPCSAMVGDLCGSIFLVSQLFTPSPRHET 300
<hr/>			
QY	301	VQDCNCISYPCHVSGHR	NAMDMWNWPTTALVVSOLLRIPOAVDDMVAGAHGVLAGLA 360
DB	301	VQDCNCISYPCHVSGHR	NAMDMWNWPTTALVVSOLLRIPOAVDDMVAGAHGVLAGLA 360
<hr/>			
QY	361	YYSMVGNWAKULI	VALLFAGVDGETHTTGTVAGHTTGFTSLFSSGASOKIQLVNTNGSW 420
DB	361	YYSMVGNWAKULI	VALLFAGVDGETHTTGTVAGHTTGFTSLFSSGASOKIQLVNTNGSW 420
<hr/>			
QY	421	HINFTALNCNDSLOTGT	GFAALFYAHKNSSGPCERMASCRPIDMFAQGWSGITTYTKNSS 480
DB	421	HINFTALNCNDSLOTGT	GFAALFYTHKNSPGCPERMASCRPIDMFAQGWSGITTYTKNSS 480
<hr/>			
QY	481	DQRYCWHYARPCCGV	PASOVCGPVYCTPSVVVGTIDBSGYPTYSWGNETDVMLN 540
DB	481	DQRYCWHYARPCCGV	PASOVCGPVYCTPSVVVGTIDBSGYPTYSWGNETDVMLN 540

541	QY	NTRPQGNWFGCTWMNS*GFTKTCGGPPCNTGGVGNRTLI	CPTDFCRKHPBATTYTKCGSG	600
541	DB	NTRPQGNWFGCTWMNS*GFTKTCGGP*CNTGGVGNRTLI	CPTDFCRKHPBATTYTKCGSG	600
601	QY	PWLTPRC:VDPYRLWHYPCTIN*TSIFKVRMVVGGVEHRLNACNWTGRCNLEDRDRS		660
601	DB	PWLTPRC:VDPYRLWHYPCTIN*TSIFKVRMVVGGVEHRLNACNWTGRCNLEDRDRS		660
661	QY	ELSP*LLSTTBEQIILPCAF*TTLPALSTGLIHLQNI*VDQVLYGVSFAVFAIKWEYIL		720
661	DB	ELSP*LLSTTBEQIILPCAF*TTLPALSTGLIHLQNI*VDQVLYGVSFAVFAIKWEYIL		720
721	QY	LIFLLIADARVCACILMMLLIAQAAEALENILVILNAA*SVAGAHGILSLVFPFCARWYIKG		780
721	DB	LIFLLIADARVCACILMMLLIAQAAEALENILVILNAA*SVAGAHGILSLVFPFCARWYIKG		780
781	QY	RLAPGAAYAFYGVWPLLILLLI*ALPRAVALOREMAASCGGAVLGLV*FTLSPPYKVFIL		840
781	DB	RLAPGAAYAFYGVWPLLILLLI*ALPRAVALOREMAASCGGAVLGLV*FTLSPPYKVFIL		840
841	QY	RLI*WMLQVFI*TRA*EAH*QWV*PPLNVRGGRDAIILITCAVHP*ELIFDITKLLAILIGPLM		900
841	DB	RLI*WMLQVFI*TRA*EAH*QWV*PPLNVRGGRDAIILITCAVHP*ELIFDITKLLAILIGPLM		900
901	QY	VLOAGHITRVPYFVRAQGLIRACMLVRKVAGGHYQVFMKLGALITGYVYNNH*LTBLRDWA		960
901	DB	VLOAGHITRVPYFVRAQGLIRACMLVRKVAGGHYQVFMKLGALITGYVYNNH*LTBLRDWA		960
961	QY	HAGLRDLAVAPVVPVFSAMETKVI*TWGADTAA*CGDIIILGLPVSARGKEIIFLGPADSLBG		1020
961	DB	HAGLRDLAVAPVVPVFSAMETKVI*TWGADTAA*CGDIIILGLPVSARGKEIIFLGPADSLBG		1020
1021	QY	QGW*LLAPITAY*SQOTRGV*LGCIITSLTGRDKNQVEGEVQV*VSTATQSP*LATCINGCVMT		1080
1021	DB	QGW*LLAPITAY*SQOTRGV*LGCIITSLTGRDKNQVEGEVQV*VSTATQSP*LATCINGCVMT		1080
1081	QY	VYHGAGSKTLAGEKGI*ITOMY*TNVDL*LVGWQAPPGARSMT*PCSCGSSDLYLV*TRHADVI		1140
1081	DB	VYHGAGSKTLAGEKGI*ITOMY*TNVDL*LVGWQAPPGARSMT*PCSCGSSDLYLV*TRHADVI		1140
1141	QY	PVRRGDSRGSLLSPRPVSYLKGSSGGPLLCP*SGHVGVVFRAAVCTRGVAKAVDFI*PVES		1200
1141	DB	PVRRGDSRGSLLSPRPVSYLKGSSGGPLLCP*SGHVGVVFRAAVCTRGVAKAVDFI*PVES		1200
1201	QY	METT*MSPVFTDNSTPPAVPQ*FOVAHLHAPT*GSGKSTKVP*AA*YAAQ*QKVL*VLPNSVAA		1260
1201	DB	METT*MSPVFTDNSTPPAVPQ*FOVAHLHAPT*GSGKSTKVP*AA*YAAQ*QKVL*VLPNSVAA		1260
1261	QY	TLGFCAYMSKAHGDINIR*TVRTITGGSITVSTY*GKFLAOGCGSGGAYDIIICDECHS		1320
1261	DB	TLGFCAYMSKAHGDINIR*TVRTITGGSITVSTY*GKFLAOGCGSGGAYDIIICDECHS		1320
1321	QY	TDS*TTILGTCVLDOA*ETAGARLV*LATAPG*SVTPHPNT*EEIGLSNNGI*PIPVGKAI		1380
1321	DB	TDS*TTILGTCVLDOA*ETAGARLV*LATAPG*SVTPHPNT*EEIGLSNNGI*PIPVGKAI		1380
1381	QY	PIEAIKGGRHILFCHSKKKCDELA*AKLITGLGLN*AVAYYRGLDVS*VITPP*IGDVVV*VATDAJ		1440
1381	DB	PIEAIKGGRHILFCHSKKKCDELA*AKLITGLGLN*AVAYYRGLDVS*VITPP*IGDVVV*VATDAJ		1440
1441	QY	MTGFTGDFSVIDCNTCV*QTVD*FSLD*PFTT*ETT*TV*PQDAVRSR*ORRGTGRGSGIYR		1500
1441	DB	MTGFTGDFSVIDCNTCV*QTVD*FSLD*PFTT*ETT*TV*PQDAVRSR*ORRGTGRGSGIYR		1500
1501	QY	FVTPGSRP*SGMFD*SSVLCB*CYDAGCAWYELT*PAETS*VRL*VAYL*NPGL*VPCODHLE*FWES		1560
1501	DB	FVTPGSRP*SGMFD*SSVLCB*CYDAGCAWYELT*PAETS*VRL*VAYL*NPGL*VPCODHLE*FWES		1560
1561	QY	VFTGLTHIDAH*FUSQ*OKAGDN*FPYLVA*QATV*CARAQAP*PPSWD*QW*CLIRL*KT*HLG		1620
1561	DB	VFTGLTHIDAH*FUSQ*OKAGDN*FPYLVA*QATV*CARAQAP*PPSWD*QW*CLIRL*KT*HLG		1620
1621	QY	PT*LLYRLGAVQNEVIL*THPITKYIN*MACNSAD*LEV*VSTWILVGGV*LAALAY*CLIT*GVS		1680

1621	Db	PTPLLYFLGAVQNEVILTHPTIYIYIACMSADLEWVTSTWVLVGVYLAALAAAYCLATTGVS	1680
1681	Qy	VIYGRILISCKPAVDPREVLYQSEDEMESCASQLPYIEGNOQLAQBFQKALGLHLOAT	1740
1681	Db	VIYGRILISCKPAVDPREVLYQSEDEMESCASQLPYIEGNOQLAQBFQKALGLHLOAT	1740
1741	Qy	KQAEAAAFVVESKWRALETFWAKHMNFISGIOYLAGLSLPGNPAIASLMAFTASITSP	1800
1741	Db	KQAEAAAFVVESKWRALETFWAKHMNFISGIOYLAGLSLPGNPAIASLMAFTASITSP	1800
1801	Qy	LTTQNTLLFNILGWSAAQLAPSAASAFVAGAGIAGAAVCSIGLGKVLVDILAGYGAVA	1860
1801	Db	LTTQNTLLFNILGWSAAQLAPSAASAFVAGAGIAGAAVCSIGLGKVLVDILAGYGAVA	1860
1861	Qy	GALVAPKVMGEVPSDEDLVNLKLPAILLSPGALVGVVCAAILLRHHVGPGEVQVWNRLLI	1920
1861	Db	GALVAPKVMGEVPSDEDLVNLKLPAILLSPGALVGVVCAAILLRHHVGPGEVQVWNRLLI	1920
1921	Qy	AFASRGNHVSPTHYVPPESDAAARVTQILSLTITQLLKRLHQMINEDCSTPCSGSMLRDV	1980
1921	Db	AFASRGNHVSPTHYVPPESDAAARVTQILSLTITQLLKRLHQMINEDCSTPCSGSMLRDV	1980
1981	Qy	WDMLCTVLTFDKWVLQSKLLPRPGVPPFLSCQRGYKGVWEGDGMQTTCPGQAIIAGHVK	2040
1981	Db	WDMLCTVLTFDKWVLQSKLLPRPGVPPFLSCQRGYKGVWEGDGMQTTCPGQAIIAGHVK	2040
2041	Qy	NGSMRIYVGPRTCSNTHGTFPIINAYTTGCTPSPAPNYSRALRVAAEFVEVTRVGDFF	2100
2041	Db	NGSMRIYVGPRTCSNTHGTFPIINAYTTGCTPSPAPNYSRALRVAAEFVEVTRVGDFF	2100
2101	Qy	YVTGMTTDNVKCPQVPAPEFFTEVDGVRLHRYPACKPLLEDVTFQVGLNQYLVGSOL	2160
2101	Db	YVTGMTTDNVKCPQVPAPEFFTEVDGVRLHRYPACKPLLEDVTFQVGLNQYLVGSOL	2160
2161	Qy	PCPEPDPVTLTSMLTDPSHITAEAKRRLARGSPPSLASSASQASGLKATCTTHHD	2220
2161	Db	PCPEPDPVTLTSMLTDPSHITAEAKRRLARGSPPSLASSASQASGLKATCTTHHD	2220
2221	Qy	SPADLLLEANLWPEQMGNIITRVESENKVILDSPEFLHAEGDDEBISVAAEILAKSRK	2280
2221	Db	SPADLLLEANLWPEQMGNIITRVESENKVILDSPEFLHAEGDDEBISVAAEILAKSRK	2280
2281	Qy	FPSALP.IWA.RPDY.NP.LLES.KW.DPY.VPV.VHG.CPL.PTKAP.TTP.PPR.KRT.VVL.TES.NV	2340
2281	Db	FPSALP.IWA.RPDY.NP.LLES.KW.DPY.VPV.VHG.CPL.PTKAP.TTP.PPR.KRT.VVL.TES.NV	2340
2341	Qy	SSALAEIATKTFPGSGGSANDVGATATLPD.LASDDG.KGSDVSZSY.SSNP.PLEGE.PGDP.L	2400
2341	Db	SSALAEIATKTFPGSGGSANDVGATATLPD.LASDDG.KGSDVSZSY.SSNP.PLEGE.PGDP.L	2400
2401	Qy	SDGSWS.TV.SBEAS.DV.VCCNS.Y.TWTGAL.IT.PCAABESKLPINPLNSL.IR.HNNM.VYATT	2460
2401	Db	SDGSWS.TV.SBEAS.DV.VCCNS.Y.TWTGAL.IT.PCAABESKLPINPLNSL.IR.HNNM.VYATT	2460
2461	Qy	SRSASLRQK.VT.FOR.LQ.VLD.DH.YRD.VL.KEM.KAKASTV.KAKLLS.IEEACKL.TP.PHS.AK.SKF	2520
2461	Db	SRSASLRQK.VT.FOR.LQ.VLD.DH.YRD.VL.KEM.KAKASTV.KAKLLS.IEEACKL.TP.PHS.AK.SKF	2520
2521	Qy	GYGAKDV.NI.LSSRAV.NH.IRS.WVED.LLBDET.PID.TT.IMAKSE.VFCVQPEXGGRK.PAR.LIV	2580
2521	Db	GYGAKDV.NI.LSSRAV.NH.IRS.WVED.LLBDET.PID.TT.IMAKSE.VFCVQPEXGGRK.PAR.LIV	2580
2581	Qy	FPDLGVRV.CEK.MAL.YDV.VSTLP.QA.VMGSS.YGFQ.SP.QRV.FEFL.VNT.WKSKC.KP.GFS.YDT	2640
2581	Db	FPDLGVRV.CEK.MAL.YDV.VSTLP.QA.VMGSS.YGFQ.SP.QRV.FEFL.VNT.WKSKC.KP.GFS.YDT	2640
2641	Qy	RCFDS.TV.TESDIR.VEESI.YQC.DLAP.EARQ.AIRSI.TBRL.YIGG.PLTNS.KGONC.YR.CRA	2700
2641	Db	RCFDS.TV.TESDIR.VEESI.YQC.DLAP.EARQ.AIRSI.TBRL.YIGG.PLTNS.KGONC.YR.CRA	2700
2701	Qy	SGVLT.TSCG.NTLT.CYL.KATAA.CRAA.KLQD.CT.MLVNG.DDLV.VIC.SAGTQEDAAALRAFTE	2760

Dd	2701	SGVLTTCGNTLTCYLKATACRAAKLQDCTMLVNGDDLVIICESAGTQDAALRAFTE	2760
Qy	2761	AMTRYAPPDPQPPEYDLLEITSCSSNSVAHDASGRKYVYILTRDPTTFLAARAWE ^{TAR}	2820
Dd	2761	AMTRYAPPDPQPPEYDLLEITSCSSNSVAHDASGRKYVYILTRDPTTFLAARAWE ^{TAR}	2820
Qy	2821	HTPINSWLGNIMWAPTLWARMLINTHFFSILLAQLEKALDCQIYGACYSIEPLDLPQ	2880
Dd	2821	HTPINSWLGNIMWAPTLWARMLINTHFFSILLAQLEKALDCQIYGACYSIEPLDLPQ	2880
Qy	2881	IIRLHGLSAFTLLHSYSPGEINRVASCLRKLGVPPPLRTWHRARSVRAKLLSQGRAATC	2940
Dd	2881	IIRLHGLSAFTLLHSYSPGEINRVASCLRKLGVPPPLRTWHRARSVRAKLLSQGRAATC	2940
Qy	2941	GRYLFNWA ^{VET} TKLKLTPIPAASQLDSGFNFVAGYSGGDIYHLSRARPRWFPCLLLLSV	3000
Dd	2941	GRYLFNWA ^{VET} TKLKLTPIPAASQLDSGFNFVAGYSGGDIYHLSRARPRWFPCLLLLSV	3000
Qy	3001	GVGIYLLPNR 3010 	
Dd	3001	GVGIYLLPNR 3010 	
RESULT 4			
ID	Q02828	PRELIMINARY; PRT; 3010 AA.	
AC	Q02828;	1996 (TrEMBLrel. 01, Created)	
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	Polyprotein precursor (Genome polyprotein).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepadnavirus.		
OX	NCBI_TaxID=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=HC-J4;		
RX	MEDLINE=92391112; PubMed=1325713;		
RA	Okamoto H., Kojima M., Okada S., Yoshizawa H., Iizuka H., Tanaka T., Muchmore E.B., Peterson D.A., Ito Y., Mishiro S.;		
RT	"Genetic drift of Hepatitis C virus during an 8.2-year infection in a rhesus chimpanzee: variability and stability."		
RL	Virology 190:894-899(1992).		
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).		
CC	EXBL; D10750; BAA01583.1; "-		
PIR	A61196; A61196.		
DR	PIR; PQ0246; PQ0246.		
DR	PIR; PS0329; PS0329.		
DR	FDB; INB4; 25-MAR-03.		
DR	GO; GO:0016031; C:integral to membrane; IEA.		
DR	GO; GO:0019028; C:viral capsid; IEA.		
DR	GO; GO:0019031; C:viral envelope; IEA.		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0008026; F:ATP dependent helicase activity; IEA.		
DR	GO; GO:0005489; F:electron transporter activity; IEA.		
DR	GO; GO:0003723; F:RNA binding; IEA.		
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.		
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.		
D3	GO; GO:0005198; F:structural molecule activity; IEA.		
D2	GO; GO:0016740; F:transferase activity; IEA.		
DR	GO; GO:0006118; F:electron transport; IEA.		
DR	GO; GO:0006508; F:proteolysis and peptidolysis; IEA.		
DR	GO; GO:0006350; F:transcription; IEA.		
DR	GO; GO:0019079; F:viral genome replication; IEA.		
DR	GO; GO:0019057; F:viral transformation; IEA.		
DR	InterPro; IPRO00003; Cys Ser trypsin.		
DR	InterPro; IPRO00345; CytC_heme_BS.		
DR	InterPro; -IPRO01410; DRAD.		
DR	InterPro; IPRO02522; HCV capsid.		

DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RGR.
DR InterPro; IPR004109; Pepsidase C29.
DR InterPro; IPR007095; RNA pol D3 PS.
DR InterPro; IPR007094; RNA pol PSvir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191 C PROTEIN.
FT CHAIN 192 383 E PROTEIN.
FT CHAIN 384 729 NS1/E2 PROTEIN.
FT CHAIN 730 1006 NS2 PROTEIN.
FT CHAIN 1007 1615 NS3 PROTEIN.
FT CHAIN 1616 2013 NS4 PROTEIN.
FT CHAIN 2014 3010 NS5 PROTEIN.
SQ SEQUENCE 3010 AA; 326600 MW; F44C7FE7D288C242 CRC64;

Query Match 99.4%; Score 15909; DB 12; Length 3010;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 2991; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSTNPKPQKTKNTNRRQDVKFPGGQIVGGVILLPRRGLGVRAATKASERSQPRG 60
DB 1 MSTNPKPQKTKNTNRRQDVKFPGGQIVGGVILLPRRGLGVRAATKASERSQPRG 60
QY 61 RKQIPKARPEGRAWAQGYFMPLYGNEGLGWAGLLSPGRSPGSGPTDPRRSNGL 120
DB 61 RKQIPKARPEGRAWAQGYFMPLYGNEGLGWAGLLSPGRSPGSGPTDPRRSNGL 120
QY 121 KVDTLTGCFADLMGYIPLVGAELGGAARALAKGVRLVEDGVNYATGNLPCCSFSIFLLA 180
DB 121 KVDTLTGCFADLMGYIPLVGAELGGAARALAKGVRLVEDGVNYATGNLPCCSFSIFLLA 180
QY 181 LLSCLTIPASAYEVRNVSGIYHVNTDCNSSIYVEAADVIMHTPGCVPCVQEGNSRRCW 240
DB 181 LLSCLTIPASAYEVRNVSGIYHVNTDCNSSIYVEAADVIMHTPGCVPCVQEGNSRRCW 240
QY 241 ALPTLAARNASVPTTIRRHVDLLVGTAAFCASMYVDLCCGSIFLVQSQFTSPRHHET 300
DB 241 ALPTLAARNASVPTTIRRHVDLLVGTAAFCASMYVDLCCGSIFLVQSQFTSPRHHET 300
QY 301 VQDNCNLIYPGHVSGHGMAMNMNNSPTTALVVSQILRIPOAVDMVAGAHGVLGLA 360
DB 301 VQDNCNLIYPGHVSGHGMAMNMNNSPTTALVVSQILRIPOAVDMVAGAHGVLGLA 360
QY 361 YYSVGNWAKVLIALLFAGVDGETHTTGRVAGHTTSGFTSLFSSGASQKIQLVNTNGSW 420
DB 361 YYSVGNWAKVLIALLFAGVDGATYTSGGVAGETTSFTSGFTSLFSSGASQKIQLVNTNGSW 420
QY 421 HINRTALNCNDSLTGTGFAALFYAHKFNSSGCPERVASCRPIDWFAQGWGPITYTKPNSS 480
DB 421 HINRTALNCNDSLTGTGFAALFYTHKFNSSGCPERVASCRPIDWFAQGWGPITYTKPNSS 480
QY 481 DQRPYCHWYAPRCGVVYASQVGCFTPTSPVVGTTDRSGVPTYSWGENETDVMLLN 540

DB 481 DQRPYCHWYAPRCGVVYASQVGCFTPTSPVVGTTDRSGVPTYSWGENETDVMLLN 540
QY 541 NTRPPQGNWFGCTWNNSTGFTKTCGGPPCNIGGVGNRTLIPTDCFRKHPEATYTKCGSG 600
DB 541 NTRPPQGNWFGCTWNNSTGFTKTCGGPPCNIGGVGNRTLIPTDCFRKHPEATYTKCGSG 600
QY 601 PWTPLRCCLVDYPYRLWHYPCTLNFSIFKVMYVGGVHRHNAACNWTGRCNLEDDRDS 660
DB 601 PWTPLRCCLVDYPYRLWHYPCTLNFSIFKVMYVGGVHRHNAACNWTGRCNLEDDRDS 660
QY 661 ELSPLLLSTTEWQILPCAFPTTLPALSTGLIHLHQNIVDVQVLYGVGSFVFAIKWEYIL 720
DB 661 ELSPLLLSTTEWQILPCAFPTTLPALSTGLIHLHQNIVDVQVLYGVGSFVFAIKWEYIL 720
QY 721 LFLFLLADARVCACILWMLLIAQAALLENLVNLAASVAGAHGILSLFVFFCAAWYIKG 780
DB 721 LFLFLLADARVCACILWMLLIAQAALLENLVNLAASVAGAHGILSLFVFFCAAWYIKG 780
QY 781 RLAPGAAYAFYGVWPLLLLLLALPPRAYALDREMAASCGGAVLVCLVLTSPYKVFLLT 840
DB 781 RLAPGAAYAFYGVWPLLLLLLALPPRAYALDREMAASCGGAVLVCLVLTSPYKVFLLT 840
QY 841 RLIMWLOYEITRAEAHMOVWVPLNVGRDAIILLTCAVHPELIFDTIKLLAILGLPLX 900
DB 841 RLIMWLOYEITRAEAHMOVWVPLNVGRDAIILLTCAVHPELIFDTIKLLAILGLPLX 900
QY 901 VLQAGITRVPYFVRAQGLIRACMLVRKVAGGHYVQVFMKLGALTGTYYNELTPLRDWA 960
DB 901 VLQAGITRVPYFVRAQGLIRACMLVRKVAGGHYVQVFMKLGALTGTYYNELTPLRDWA 960
QY 961 HAGLDLAVAVEPVVFSAMETKVITWADTAACGDIILGLPVSARRGKEIFLGPADSLEG 1020
DB 961 HAGLDLAVAVEPVVFSAMETKVITWADTAACGDIILGLPVSARRGKEIFLGPADSLEG 1020
QY 1021 QGWRLLAPITAYSQQTRGVLCIITSLTGRDNQVGEVQVNSTATQSLATCINGVCWT 1080
DB 1021 QGWRLLAPITAYSQQTRGVLCIITSLTGRDNQVGEVQVNSTATQSLATCINGVCWT 1080
QY 1081 VYHAGSKXTLAGPKGPIQMTYTNVDLDLVGWAQPPGARSMTPCSCGSDLYLVTRHADVI 1140
DB 1081 VYHAGSKXTLAGPKGPIQMTYTNVDLDLVGWAQPPGARSMTPCSCGSDLYLVTRHADVI 1140
QY 1141 PYRRGDSRGSLLSPRPVSYLKGSGGELLCPSGHVGVFRAAVCTRGVAKAVDIPVES 1200
DB 1141 PYRRGDSRGSLLSPRPVSYLKGSGGELLCPSGHVGVFRAAVCTRGVAKAVDIPVES 1200
QY 1201 METTWRSPVFTDNSTPPAVPOTFQVAHLHAPTGSCKSTKVPAAVAAQGVKVLNPSVAA 1260
DB 1201 METTWRSPVFTDNSTPPAVPOTFQVAHLHAPTGSCKSTKVPAAVAAQGVKVLNPSVAA 1260
QY 1261 TLGFGAYMSKAHGDIPNIRGTGRTITGSGITSTYTGKPLADGGCGSGAYDIIICDECHS 1320
DB 1261 TLGFGAYMSKAHGDIPNIRGTGRTITGSGITSTYTGKPLADGGCGSGAYDIIICDECHS 1320
QY 1321 TQSTTILGIGTVLDQAEAGARLVLATATPGSVTVPHNIEETGLSNNGSIPIYGGAI 1380
DB 1321 TQSTTILGIGTVLDQAEAGARLVLATATPGSVTVPHNIEETGLSNNGSIPIYGGAI 1380
QY 1381 PTEALKGGRHLIFCHSKKCDLAAKLTLGLGNAYVYRGLDVSVIPIPTGDVVVATDAL 1440
DB 1381 PTEALKGGRHLIFCHSKKCDLAAKLTLGLGNAYVYRGLDVSVIPIPTGDVVVATDAL 1440
QY 1441 MTGFTGDPDSIDNCNCTVQTVDFSLDPTFTTTTTPQDAVSRQRGRGTGRSGIYR 1500
DB 1441 MTGFTGDPDSIDNCNCTVQTVDFSLDPTFTTTTTPQDAVSRQRGRGTGRSGIYR 1500
QY 1501 FVTGPRPSGMPDSSVLCEDYDAGAWYELTAPETSVRLRAYLNTPGLPVCCDHLEFWS 1560
DB 1501 FVTGPRPSGMPDSSVLCEDYDAGAWYELTAPETSVRLRAYLNTPGLPVCCDHLEFWS 1560
QY 1561 VFTGTTHIDAHFLSOTKQAGDNFPYLVAYQATVCARAQAPPSPQWQMKLIRLKPTLHG 1620
DB 1561 VFTGTTHIDAHFLSOTKQAGDNFPYLVAYQATVCARAQAPPSPQWQMKLIRLKPTLHG 1620

QY 1621 PTPLVRLGAVQNEVLTTHPIITKYIMACMSADLEVVTSTWLVGGVLAALAAVCLTTGVS 1680
DB 1621 PTPLVRLGAVQNEVLTTHPIITKYIMACMSADLEVVTSTWLVGGVLAALAAVCLTTGVS 1680
QY 1681 VIVGRILSGKPAVDPREVLYQEFDEMEECASQLPYIEQGMQABQFKQKALGLLQAT 1740
DB 1681 VIVGRILSGKPAVDPREVLYQEFDEMEECASQLPYIEQGMQABQFKQKALGLLQAT 1740
QY 1741 KQREAAAPVVEKSWRALETFWAKHMNFISGIIYLAGLSTLPCNPAPAIASMAPTASITSP 1800
DB 1741 KQREAAAPVVEKSWRALETFWAKHMNFISGIIYLAGLSTLPCNPAPAIASMAPTASITSP 1800
QY 1801 LTTQNTLLFNILGGVAAQALAPSAASAFVAGIAGAAVSGIGLKVLDVILAGYGAVA 1860
DB 1801 LTTQNTLLFNILGGVAAQALAPSAASAFVAGIAGAAVSGIGLKVLDVILAGYGAVA 1860
QY 1861 GALVAFKMSGEVSTEDLVNLLPALISPGALVGVVCAILRRHYGPGEGAVQVWNRLLI 1920
DB 1861 GALVAFKMSGEVSTEDLVNLLPALISPGALVGVVCAILRRHYGPGEGAVQVWNRLLI 1920
QY 1921 AFASRGHVSPTHYVESDAAARVTOILLSLITITOLLKRLHWINEDCSTPCSGSMLRDV 1980
DB 1921 AFASRGHVSPTHYVESDAAARVTOILLSLITITOLLKRLHWINEDCSTPCSGSMLRDV 1980
QY 1981 WDMICTVLTDFTKWLQSKLLPRLPGVDFLSCQRYKGVWGRGDIMOTTCPGQAQIAGHVX 2040
DB 1981 WDMICTVLTDFTKWLQSKLLPRLPGVDFLSCQRYKGVWGRGDIMOTTCPGQAQIAGHVX 2040
QY 2041 NGSMRIVGPRTCNTWHGTTPINAYTTGCTPSPAPNYSALMRVAEEYVEVTRVGDFF 2100
DB 2041 NGSMRIVGPRTCNTWHGTTPINAYTTGCTPSPAPNYSALMRVAEEYVEVTRVGDFF 2100
QY 2101 YVTGMTTDNVKCPQVPAPEFFTEVDGRLRHYPACKPILREDTTFQVGNQYLVGSOL 2160
DB 2101 YVTGMTTDNVKCPQVPAPEFFTEVDGRLRHYPACKPILREDTTFQVGNQYLVGSOL 2160
QY 2161 PCPEPDPVTLTSLMTPDPSHITAEAKRLARGSPPSLASSASQLSAPLSKATCTTHD 2220
DB 2161 PCPEPDPVTLTSLMTPDPSHITAEAKRLARGSPPSLASSASQLSAPLSKATCTTHD 2220
QY 2221 SPDALEALLMRQEMGNITRVESKNVILDSFEPFLHAGDEREISVAAEILRSKR 2280
DB 2221 SPDALEALLMRQEMGNITRVESKNVILDSFEPFLHAGDEREISVAAEILRSKR 2280
QY 2281 FPSALPIWAPDYNPPLLESWKDDYVPPVYVHGCLPPTKAPPTPPRRKRTVVLTSSNV 2340
DB 2281 FPSALPIWAPDYNPPLLESWKDDYVPPVYVHGCLPPTKAPPTPPRRKRTVVLTSSNV 2340
QY 2341 SSALAEALATKTFGSSGSAVDSGTATALPDLASDDGDKGSDVESYSSMPPLEGEPGPDL 2400
DB 2341 SSALAEALATKTFGSSGSAVDSGTATALPDLASDDGDKGSDVESYSSMPPLEGEPGPDL 2400
QY 2401 SDGWSVTVEASEBVDVCCMSYTWTCALITPCAABESKLPINPLNSLLPHNNVYATT 2460
DB 2401 SDGWSVTVEASEBVDVCCMSYTWTCALITPCAABESKLPINPLNSLLPHNNVYATT 2460
QY 2461 SRSASLRQKVTDFRLQVLDHVDLKMKAASKASTVAKALLSTEEACKLTTPHSASKF 2520
DB 2461 SRSASLRQKVTDFRLQVLDHVDLKMKAASKASTVAKALLSTEEACKLTTPHSASKF 2520
QY 2521 GYGAKDVNLSSRAVNHRTSVWEDLLEDTETPIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
DB 2521 GYGAKDVNLSSRAVNHRTSVWEDLLEDTETPIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
QY 2581 FPDJGVRCVCKXALYDVVSTLPOAVMGSSYGFQYSPKORVFLVNTWKSXKCPMGFSYDT 2640
DB 2581 FPDJGVRCVCKXALYDVVSTLPOAVMGSSYGFQYSPKORVFLVNTWKSXKCPMGFSYDT 2640
QY 2641 RCFDSTVTVESIRVEESIQCCDLAPARQAIRSLTERLYTGGPLTNSKGQNCVGRRA 2700
DB 2641 RCFDSTVTVESIRVEESIQCCDLAPARQAIRSLTERLYTGGPLTNSKGQNCVGRRA 2700

QY 2701 SGVLTTSCTNTLTCYLKATAACRAAKLQDCTMLVNGDDLVTCEBSAGTOEDAAALPAFTE 2760
DB 2701 SGVLTTSCTNTLTCYLKATAACRAAKLQDCTMLVNGDDLVTCEBSAGTOEDAAALPAFTE 2760
QY 2761 AMTRYSAAPGDPQPEYDLELITSCSSNVSAHDASGRVYVLTDRDPTPLAPAAWETAR 2820
DB 2761 AMTRYSAAPGDPQPEYDLELITSCSSNVSAHDASGRVYVLTDRDPTPLAPAAWETAR 2820
QY 2821 HTPINSLWLNITMYAFTLWARMLMTHFFSILLAQOELEKALDCQIYGACYSIEPLDLPQ 2880
DB 2821 HTPINSLWLNITMYAFTLWARMLMTHFFSILLAQOELEKALDCQIYGACYSIEPLDLPQ 2880
QY 2881 IITERLHGLSAFTLHYSYSGEINRVASCLRKLVGPPLRTWRHARSVRAKLLSQGGAATC 2940
DB 2881 IITERLHGLSAFTLHYSYSGEINRVASCLRKLVGPPLRTWRHARSVRAKLLSQGGAATC 2940
QY 2941 GSYLFNVAWRTKLKUTPIPAASQDLDSGFVAGYSGGDIYHLSRARPRWFPLCLILLFV 3000
DB 2941 GSYLFNVAWRTKLKUTPIPAASQDLDSGFVAGYSGGDIYHLSRARPRWFPLCLILLFV 3000
QY 3001 GVGIVYLLPNR 3010
DB 3001 GVGIVYLLPNR 3010
RESULT 5
Q52972 PRELIMINARY; PRT; 3010 AA.
ID Q52972; AC Q52972; DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J4;
RC MEDLINE=98240944; PubMed=9581788;
RA Yanagi M., St. Claire M., Shapiro M., Emerson S.U., Purcell R.H.,
RA Bukh J.;
RT "Transcripts of a chimeric clone of hepatitis C virus genotype 1b are
RT infectious in vivo."
RL Virology 244:161-172(1998).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF054250; AAC15725.1; -;
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PS0329; PS0329.
DR HSP; P28663; Iuxp.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.

DR	InterPro; IPR002519; HCV_env.
DR	InterPro; IPR002531; HCV_NSI.
DR	InterPro; IPR002518; HCV_NS2.
DR	InterPro; IPR000745; HCV_NS4a.
DR	InterPro; IPR001490; HCV_NS4b.
DR	InterPro; IPR002868; HCV_NS5a.
DR	InterPro; IPR002166; HCV_RdRp.
DR	InterPro; IPR004109; Peptidase_C29.
DR	InterPro; IPR007095; RNA_pol_DS_PS.
DR	InterPro; IPR007094; RNA_pol_PSVir.
DR	Pfam; PF01543; HCV_capsig; 1.
DR	Pfam; PF01542; HCV_core; 1.
DR	Pfam; PF01539; HCV_env; 1.
DR	Pfam; PF01560; HCV_NSI; 1.
DR	Pfam; PF01338; HCV_NS2; 1.
DR	Pfam; PF02907; HCV_NS3; 1.
DR	Pfam; PF01006; HCV_NS4a; 1.
DR	Pfam; PF01031; HCV_NS4b; 1.
DR	Pfam; PF01506; HCV_NS5a; 1.
DR	Pfam; PF00998; Viral_RdRp; 1.
DR	ProDom; PD186062; HCV_NSI; 1.
DR	SMART; SM00487; DEXDc; 1.
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW	Polyprotein; RNA-directed RNA polymerase; Transfrase; Transmembrane.
SQ	SEQUENCE 3010 AA; 326729 MW; A14CA74B4197B5 CRC64;
Query Match 99.2%; Score 15887; DB 12; Length 3010;	
Best Local Similarity 99.2%; Pred. No. 0;	
Matches 2987; Conservative 2; Mismatches 21; Indels 0; Gaps 0;	
QY	1 MSTNPKPQRKTENTNRPPDVFPFGGQIVGGVYLLPRGPRLGVATRKASRSQPRG 60
DB	1 MSTNPKPQRKTENTNRPPDVFPFGGQIVGGVYLLPRGPRLGVATRXKXSERSQPRG 60
QY	61 RRQPIPKAREPEGRAWAOCYPPLYLENEGILGNAGWLLSPRGSPSGPTDPRRSNLG 120
DB	61 RRQPIPKARXPEGRAWAOCQTPPLYLENEGILGNAGWLLSPRGSPSGPTDPRRSNLG 120
QY	121 KVIDTLTCGFADLNGYIPLVGAPLGGAARALAHGVRLVEDGWNVATGNLPCCSFIFLLA 180
DB	121 KVIDTLTCGFADLNGYIPLVGAPLGGAARALAHGVRLVEDGDVVATGNLPCCSFIFLLA 180
QY	181 LLSCLTIIPASAYEVRNVSGIIHYVTNDSCNSIIVEAADVIHMTPCVCPCVEGNSRCRW 240
DB	181 LLSCLTIIPASAYEVRNVSGIIHYVTNDSCNSIIVEAADVIHMTPCVCPCVEGNSRCRW 240
QY	241 ALPTETLAARNASVPTTIRRHVDLLVGTAAFCFSAMYVGLDCGSIFLVSQLFTFSPRRHET 300
DB	241 ALPTETLAARNASVPTTIRRHVDLLVGTAAFCFSAMYVGLDCGSIFLVSQLFTFSPRRHET 300
QY	301 VQDCNCISYPGHVSGERMAMDMNNWSPPTALVVSOLLRIIPAQVVDVMVAGAHWGVLAGIA 360
DB	301 VQDCNCISYPGHVSGERMAMDMNNWSPPTALVVSOLLRIIPAQVVDVMVAGAHWGVLAGIA 360
QY	361 YYSMVGNWAKVLIIVALLFAGVDGTHTTGTVAGHTTSGETSLFSSCASOKQLVNNTGWS 420
DB	361 YYSMVGNWAKVLLIIVALLFAGVDGTHTYIXXVKGHTTSGETSLFSSCASQKLQNVTNGSW 420
QY	421 HINRTALNCDSLQTGFFAALFYAFKFNSSCPEMASCRPIDWFPAQGWPITYITKPNSX 480
DB	421 HINRTALNCDSLQTGFKAALFYAHKTNSSCPEMASCRPIDWFXQGWPIITYITKPNSX 480
QY	481 DQRPYCHYAAPPGVPASPQVCGPVICFTSPVVVGTIDRSGVDPTYSGENETDWMLLN 540
DB	481 DQRPYCHYAAPPCGXPFASQVCGPVICFTSPVVVGTIDRSGVDPTYSGENETDWMLLN 540
QY	541 NTRPEQGNWFCTWNNSGTFTKCGPPCNTGGVGNRTLICTDCKFKHPHEATYTKCGSG 600
DB	541 NTRPEQGNWFCTWNNSGTFTKCGPPONTGGVGNRTLICTDCKFKHPHEATYTKCGSG 600
QY	601 PWTTPRCLVDYPIRWHYPCTLNIFSIFKRVYVVGVEHLNAACNWTRGERCNLDRDS 660
DB	601 PWTTPRCLVDYPIRWHYPCTLNIFSIFKRVYVVGVEHLNAACNWTRGERCNLDRDS 660

Qy	661	ELSP	LL	STTE	WO	IL	PC	AT	TL	PA	LS	TG	CI	JHL	HQ	NI	VD	QV	LY	VG	SA	VF	AI	KW	EY	IL	720																												
Db	661	ELSP	LL	STTE	WO	IL	PC	AT	TL	PA	LS	TG	CI	JHL	HQ	NI	VD	QV	LY	VG	SA	VF	AI	KW	EY	IL	720																												
Qy	721	LLFL	LL	ADAR	VC	AC	IL	WM	ML	LI	QA	EA	AL	EN	LV	LN	AA	SV	AG	HG	IL	SL	FL	VF	CA	AW	Y	KG	780																										
Db	721	LLFL	LL	ADAR	VC	AC	IL	WM	ML	LI	QA	EA	AL	EN	LV	LN	AA	SV	AG	HG	IL	SL	FL	VF	CA	AW	Y	KG	780																										
Qy	781	RLAP	GA	AY	AF	Y	GW	PL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	840																											
Db	781	RLAP	GA	AY	AF	Y	GW	PL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	LL	840																											
Qy	841	RLI	WM	LQ	FI	TR	AE	AH	MO	WV	PL	NV	RGR	DA	I	LL	T	CA	VE	HE	L	I	F	D	T	K	LL	AI	LG	PL	900																								
Db	841	RLI	WM	LQ	FI	TR	AE	AH	MO	WV	PL	NV	RGR	DA	I	LL	T	CA	VE	HE	L	I	F	D	T	K	LL	AI	LG	PL	900																								
Qy	901	VLQ	AG	IT	RP	Y	FR	AG	L	IR	AC	ML	VR	K	AG	H	Y	Q	M	F	M	K	L	G	A	L	T	G	Y	V	N	HL	T	PL	D	WA	960																		
Db	901	VLQ	AG	IT	RP	Y	FR	AG	L	IR	AC	ML	VR	K	AG	H	Y	Q	M	F	M	K	L	G	A	L	T	G	Y	V	N	HL	T	PL	D	WA	960																		
Qy	961	HAG	LR	D	L	A	V	AE	P	V	FS	AE	T	K	VI	T	W	G	A	D	T	A	C	G	I	L	L	G	L	P	S	A	R	G	K	E	I	F	G	P	A	D	S	L	E	G	1020								
Db	961	HAG	LR	D	L	A	V	AE	P	V	FS	AE	T	K	VI	T	W	G	A	D	T	A	C	G	I	L	L	G	L	P	S	A	R	G	K	E	I	F	G	P	A	D	S	L	E	G	1020								
Qy	1021	QGW	RL	L	A	P	I	T	A	S	Q	O	T	R	G	V	L	G	I	I	T	S	L	T	R	D	K	N	O	V	E	Q	V	Y	T	S	T	A	S	F	L	A	C	I	N	G	V	C	W	T	1080				
Db	1021	QGW	RL	L	A	P	I	T	A	S	Q	O	T	R	G	V	L	G	I	I	T	S	L	T	R	D	K	N	O	V	E	Q	V	Y	T	S	T	A	S	F	L	A	C	I	N	G	V	C	W	T	1080				
Qy	1081	VYH	AG	S	K	T	L	A	G	P	K	S	I	T	Q	M	Y	T	N	V	D	L	D	L	Y	G	W	A	P	P	G	A	R	S	M	T	P	C	S	G	S	S	D	L	Y	I	V	T	R	H	A	D	V	I	1140
Db	1081	VYH	AG	S	K	T	L	A	G	P	K	S	I	T	Q	M	Y	T	N	V	D	L	D	L	Y	G	W	A	P	P	G	A	R	S	M	T	P	C	S	G	S	S	D	L	Y	I	V	T	R	H	A	D	V	I	1140
Qy	1141	PV	R	R	D	S	R	G	S	L	L	S	P	R	P	S	Y	L	K	G	S	G	P	L	C	P	S	H	V	G	V	F	R	A	A	V	C	H	R	G	V	A	K	A	V	D	F	I	P	V	S	1200			
Db	1141	PV	R	R	D	S	R	G	S	L	L	S	P	R	P	S	Y	L	K	G	S	G	P	L	C	P	S	H	V	G	V	F	R	A	A	V																			

QY 1741 KQAEAAAPVVEGSKWRALETFWAKHMWNFTSGIOYLAGLSTLPGNPAIASMAFTASITSP 1800
DB 1741 KQAEAAAPVVEGSKWRALETFWAKHMWNFTSGIOYLAGLSTLPGNPAIASMAFTASITSP 1800
QY 1801 LTTQNTLLFNILGCVAAQLAPSAASAFVGAGIAGAVSGTGLKVLVDILLAGYGAGVA 1860
DB 1801 LTTQNTLLFNILGCVAAQLAPSAASAFVGAGIAGAVSGTGLKVLVDILLAGYGAGVA 1860
QY 1861 GALVAEKVMSGEVPSFEDLVNLLPAILSPGALVVGVCYCAILRRHVGPGEGAVQWNNRLI 1920
DB 1861 GALVAEKVMSGEVPSFEDLVNLLPAILSPGALVVGVCYCAILRRHVGPGEGAVQWNNRLI 1920
QY 1921 AFASRGNHVSPTHYVPESDAARVTOQLSSITITOLLKELHOWINEDCSTPCSSGMLRDV 1980
DB 1921 AFASRGNHVSPTHYVPESDAARVTOQLSSITITOLLKELHOWINEDCSTPCSSGMLRDV 1980
QY 1981 WDMICTVLDFTKWLQSKLLPRLPVGFPLSCQGYKGVWREGDGMQITPCGQAIAGHYK 2040
DB 1981 WDMICTVLDFTKWLQSKLLPRLPVGFPLSCQGYKGVWREGDGMQITPCGQAIAGHYK 2040
QY 2041 NGSMTIVGPRCTCNTHGTFPPINAYTTGCTPSPAPNYSRALRVAAEEVVEVTRVGDFF 2100
DB 2041 NGSMTIVGPRCTCNTHGTFPPINAYTTGCTPSPAPNYSRALRVAAEEVVEVTRVGDFF 2100
QY 2101 YVTGMTTDNVKPCQVPAPEFFTEVDGVRHRYAPACKPLLRDVTDFQVGLNQYLVGSQL 2160
DB 2101 YVTGMTTDNVKPCQVPAPEFFTEVDGVRHRYAPACKPLLRDVTDFQVGLNQYLVGSQL 2160
QY 2161 PCPEPDVTVLTSMLTDPSSHITATKARRLARGSPPSLASSASQISAPSLKATCTTHD 2220
DB 2161 PCPEPDVTVLTSMLTDPSSHITATKARRLARGSPPSLASSASQISAPSLKATCTTHD 2220
QY 2221 SPADLIEANLLWRQEMCGNITRVESNKVVILDSFEPHABGDERISVAAAILRKSX 2280
DB 2221 SPADLIEANLLWRQEMCGNITRVESNKVVILDSFEPHABGDERISVAAAILRKSX 2280
QY 2281 FPSALPIWARPDYNNPPLLESKDPDYVPPVVGCPPLPPTKAPPIPPRRKRTVVLTESNV 2340
DB 2281 FPSALPIWARPDYNNPPLLESKDPDYVPPVVGCPPLPPTKAPPIPPRRKRTVVLTESNV 2340
QY 2341 SSALAEIATKTFSSGSSAVDSGTATAPDLASDDGDKGSDVESYSNMPLSEPGPDL 2400
DB 2341 SSALAEIATKTFSSGSSAVDSGTATAPDLASDDGDKGSDVESYSNMPLSEPGPDL 2400
QY 2401 SDGSWTSVSEASDVVCCSMTWTGALTTPCAAEBSKLPINPLSNLLRHNMVYATT 2460
DB 2401 SDGSWTSVSEASDVVCCSMTWTGALTTPCAAEBSKLPINPLSNLLRHNMVYATT 2460
QY 2461 SRGASLQKKVTFDLRLQVLDHRYDLVKEMKAKASTVKAKLLSTEEACKLTPPHSAKSF 2520
DB 2461 SRGASLQKKVTFDLRLQVLDHRYDLVKEMKAKASTVKAKLLSTEEACKLTPPHSAKSF 2520
QY 2521 GYGAKOVNLSRAVNHRSWEDLLEDTETPIDTTIMAKSEFCVQPEKGRKPAELIV 2580
DB 2521 GYGAKOVNLSRAVNHRSWEDLLEDTETPIDTTIMAKSEFCVQPEKGRKPAELIV 2580
QY 2581 FPDGLGVCEKMAIYDVVSTLPOAVMGSSYGFQSPQRVEFLVNTWKSXKCPMGFSYDT 2640
DB 2581 FPDGLGVCEKMAIYDVVSTLPOAVMGSSYGFQSPQRVEFLVNTWKSXKCPMGFSYDT 2640
QY 2641 RCFDSTVTSBIRVESIYCCDIAPBARQAIRSLTERLYIGGFLTKNSKGNCCYRCRA 2700
DB 2641 RCFDSTVTSBIRVESIYCCDIAPBARQAIRSLTERLYIGGFLTKNSKGNCCYRCRA 2700
QY 2701 SGVLTTCGNTLTCLYKATACRAKLODCTMLVNGDDLVCISAGTOEDAAALAPTE 2760
DB 2701 SGVLTTCGNTLTCLYKATACRAKLODCTMLVNGDDLVCISAGTOEDAAALAPTE 2760
QY 2761 AMTRYAPPGPPOPEVDLELITSCSSNVSVAHASGKRVYVLTPTTTPPLARAAWETAR 2820
DB 2761 AMTRYAPPGPPOPEVDLELITSCSSNVSVAHASGKRVYVLTPTTTPPLARAAWETAR 2820
QY 2821 HTPINSMIGNIINVAPTILWARMILMTHFFSILLAOEQLKALDCQIYGACYSIEPLDLPQ 2880

DB 2821 HTPINSMIGNIINVAPTILWARMILMTHFFSILLAOEQLKALDCQIYGACYSIEPLDLPQ 2880
QY 2881 IIERLHGUSAFTHSYSGEINRVASCLRKLGVPPLRTWRHRARSVRKLLSQGGRAATC 2940
DB 2881 IIERLHGUSAFTHSYSGEINRVASCLRKLGVPPLRTWRHRARSVRKLLSQGGRAATC 2940
QY 2941 GRYLENWAVRTKLTPIPAASQDLDSGWFVAGYSGGDIYHLSRARPRWFFLCILLASV 3000
DB 2941 GRYLENWAVRTKLTPIPAASQDLDSGWFVAGYSGGDIYHLSRARPRWFFLCILLASV 3000
QY 3001 GVGIVLLPNR 3010
DB 3001 GVGIVLLPNR 3010
RESULT 6
Q02829 PRELIMINARY; PRT: 3010 AA.
AC Q02829;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polypeptide precursor (Genome polyprotein).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_TaxID=111103;
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J4;
RX MEDLINE=92391112; PubMed=1355713;
RA Okamoto H., Kojima M., Okada S., Yoshizawa H., Iizuka H., Tanaka T.,
RA Muchmore E.B., Peterson D.A., Ito Y., Mishihiro S.;
RT "Genetic drift of Hepatitis C virus during an 8.2-year infection in a
RT chimpanzee: variability and stability.";
RL Virology 190:834-899(1992).
CC -I- SUBUNIT; THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL; D13558; BAA02756.1; -.
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PS0329; PS0329.
DR HSSP; P26663; IUXP.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006118; F:electron transport; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.

DR InterPro: IPR002166; HCV RGRP.
DR InterPro: IPR004109; Pepsidase C29.
DR InterPro: IPR007095; RNA pol_DG_PS.
DR InterPro: IPR007094; RNA pol_PSVir.
DR Pfam: PF01543; HCV capsid; 1.
DR Pfam: PF01542; HCV core; 1.
DR Pfam: PF01539; HCV env; 1.
DR Pfam: PF01560; HCV NS1; 1.
DR Pfam: PF01538; HCV NS2; 1.
DR Pfam: PF02907; HCV NS3; 1.
DR Pfam: PF01006; HCV NS4a; 1.
DR Pfam: PF01001; HCV NS4b; 1.
DR Pfam: PF01506; HCV NS5a; 1.
DR Pfam: PF00398; Viral RdRp; 1.
DR ProDom: PD186062; HCV NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME C; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191 C PROTEIN.
FT CHAIN 192 383 E PROTEIN.
FT CHAIN 384 729 NS1/E2 PROTEIN.
FT CHAIN 730 1006 NS2 PROTEIN.
FT CHAIN 1007 1615 NS3 PROTEIN.
FT CHAIN 1616 2013 NS4 PROTEIN.
FT CHAIN 2014 3010 NS5 PROTEIN.
SQ SEQUENCE 3010 AA; 326957 MW; 93D465526F3EADF8 CRC64;

Query Match 98.5%; Score 15772; DB 12; Length 3010;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 2959; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

QY 1 MSINPKPQKTKNTNRNPDQVKFPGGGQIVGGVILLPRGPRGLGVRATRKASERSQPRG 60
DB 1 MSINPKPQKTKNTNRNPDQVKFPGGGQIVGGVILLPRGPRGLGVRATRKASERSQPRG 60
QY 61 RQPIKARPRGAWAQGYWPWLYGNBGLWAGWLLSPRGRSPWGPTDPRRSRLG 120
DB 61 RQPIKARPRGAWAQGYWPWLYGNBGLWAGWLLSPRGRSPWGPTDPRRSRLG 120
QY 121 KYVDTLTGCFADLMGYIPVAGPLGGAARALAHGVLEVDGNYATGNLPGCSFSIFLLA 180
DB 121 KYVDTLTGCFADLMGYIPVAGPLGGAARALAHGVLEVDGNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTIPASAYEVNRVNSGIYHVTNDCSNSIYVEAADVIMHTPGCVPCVQGNSSRCWV 240
DB 181 LLSCLTIPASAYEVNRVNSGIYHVTNDCSNSIYVEAADVIMHTPGCVPCVQGNSSRCWV 240
QY 241 ALTPTLAARNASVPTTIRRHVDLLNGTAAPCSAMYVGLCGSIFLVSQLFTFSPRRHET 300
DB 241 ALTPTLAARNASVPTTIRRHVDLLNGTAAPCSAMYVGLCGSIFLVSQLFTFSPRRHET 300
QY 301 VQDCNCSIPGHVSGRHMAWMMWNSPTTALVVSOLLRAIPQAVDMVAGAHGVLGAGLA 360
DB 301 VQDCNCSIPGHVSGRHMAWMMWNSPTTALVVSOLLRAIPQAVDMVAGAHGVLGAGLA 360
QY 361 YYSWGNWAKVLLTVALFAGVDGETHTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420
DB 361 YYSWGNWAKVLLTVALFAGVDGETHTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420
QY 421 HINRTALNCNDSIQTGFPAALFVAHKENSSGCPERMASCRPIDWFAQGWGPIYTYTPENSS 480
DB 421 HINRTALNCNDSIQTGFPAALFVAHKENSSGCPERMASCRPIDWFAQGWGPIYTYTPENSS 480
QY 481 DQRPYCHWYAPRCGVVPASQVGVVYCFPTSPVWGTTRDSGVPTVSWGENETDVMLLN 540
DB 481 DQRPYCHWYAPRCGVVPASQVGVVYCFPTSPVWGTTRDSGVPTVSWGENETDVMLLN 540
QY 541 NTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGNRTLICTDPCRKHPEATYTKCGSG 600
DB 541 NTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGNRTLICTDPCRKHPEATYTKCGSG 600
QY 601 PWLTFRCLVDYPYRLWHYPTCLTFPSIFKRVYVGVGVEHRLNAACNWTGRGRCNLEDRDRS 660

DB 601 PWLTFRCLVDYPYRLWHYPTCLTFPSIFKRVYVGVGVEHRLNAACNWTGRGRCNLEDRDRS 660
QY 661 ELSPLLSTTEWQILPCAFITLPAISTGLIHLHONIVDVQVLYGVGSFAFVSFAIKWEYIL 720
DB 661 ELSPLLSTTEWQILPCAFITLPAISTGLIHLHONIVDVQVLYGVGSFAFVSFAIKWEYIL 720
QY 721 LFLFLLDARVCACIWMMLLIQAQAALENLVNLAAASVAGAHGILSFLVFFCAAWTIKG 780
DB 721 LFLFLLDARVCACIWMMLLIQAQAALENLVNLAAASVAGAHGILSFLVFFCAAWTIKG 780
QY 781 RLAPCAAYAFYGVWVPLLIILALPPRAYLDRENAASCGGAVLYGLVLTLPYKYVPLT 840
DB 781 RLAPCAAYAFYGVWVPLLIILALPPRAYLDRENAASCGGAVLYGLVLTLPYKYVPLT 840
QY 841 RLIMWLOVFTTRAEAHMQVWPPLNVNRRDAIILLTCAVHPELIFDITKLLAILGFLM 900
DB 841 RLIMWLOVFTTRAEAHMQVWPPLNVNRRDAIILLTCAVHPELIFDITKLLAILGFLM 900
QY 901 VLQAGITRVYFVRAQGLIRACMLVRKVAGHYVQVMFMKLGALTGYVYVYVNHLPDRWA 960
DB 901 VLQAGITRVYFVRAQGLIRACMLVRKVAGHYVQVMFMKLGALTGYVYVYVNHLPDRWA 960
QY 961 HAGLRDLAVAVEPVVFSAMETKVITWGAADTAACGDIILGLPVSARRGKEIFLGADSLG 1020
DB 961 HAGLRDLAVAVEPVVFSAMETKVITWGAADTAACGDIILGLPVSARRGKEIFLGADSLG 1020
QY 1021 QGWRLLAPITAYSQOTRGVLCIITSLTGRDNKNOVEGEVQVYVSTATOSFLATCINGVCWT 1080
DB 1021 QGWRLLAPITAYSQOTRGVLCIITSLTGRDNKNOVEGEVQVYVSTATOSFLATCINGVCWT 1080
QY 1081 VYHAGSKTLAGPKGPIQMTNVNVDLIVGQAAAPGARSKTPCSCGSDLLVYVTRHADVI 1140
DB 1081 VYHAGSKTLAGPKGPIQMTNVNVDLIVGQAAAPGARSKTPCSCGSDLLVYVTRHADVI 1140
QY 1141 PVRRGDSRGLLSRPVSYLKGSGGGLLCPSGHVGVGVRAAAYCTRGVAKAVDFIPVES 1200
DB 1141 PVRRGDSRGLLSRPVSYLKGSGGGLLCPSGHVGVGVRAAAYCTRGVAKAVDFIPVES 1200
QY 1201 METNRSVPFTDNTSPPPAVPQTFQVAHLFAPTGSGKSTKVZAAAYAAQYKVLVLPNSVAA 1260
DB 1201 METNRSVPFTDNTSPPPAVPQTFQVAHLFAPTGSGKSTKVZAAAYAAQYKVLVLPNSVAA 1260
QY 1261 TLGFGAYSKAHGIDPNIRTGVRTITGSGSTYTYGKFLADGCGSGAYDIIICDECHS 1320
DB 1261 TLGFGAYSKAHGIDPNIRTGVRTITGSGSTYTYGKFLADGCGSGAYDIIICDECHS 1320
QY 1321 TDSITTLIGTGLVLDQAETAGARLVVLTATPPGSVTVPHNIEETCLSNNGEIPFYGKAI 1380
DB 1321 TDSITTLIGTGLVLDQAETAGARLVVLTATPPGSVTVPHNIEETCLSNNGEIPFYGKAI 1380
QY 1381 PIEAIKGGRHILFCHSKKCDDELAALKTGLGLNNAVAYYRGDLVDVSIPIGDVVVATDAL 1440
DB 1381 PIEAIKGGRHILFCHSKKCDDELAALKTGLGLNNAVAYYRGDLVDVSIPIGDVVVATDAL 1440
QY 1441 MTGFTGDPDSVTDNCTCYTOTVDPSSLDTFTTETTVPDQAVSRORRGTGRSGGIYR 1500
DB 1441 MTGFTGDPDSVTDNCTCYTOTVDPSSLDTFTTETTVPDQAVSRORRGTGRSGGIYR 1500
QY 1501 FVTGPERSGMFDSSVLCBCYDAGCAWYELTPAETSVRILRAYINTPLPVCDHLEFWES 1560
DB 1501 FVTGPERSGMFDSSVLCBCYDAGCAWYELTPAETSVRILRAYINTPLPVCDHLEFWES 1560
QY 1561 VFTGLTHIDAHFLSQTKQAGDNFPYLVAQATVCARAAPPSPQMWKCLIRLKPTELHG 1620
DB 1561 VFTGLTHIDAHFLSQTKQAGDNFPYLVAQATVCARAAPPSPQMWKCLIRLKPTELHG 1620
QY 1621 PTPELYRLGAVONEVILLTHPIITKYIMACMSADLEVYVTSWTWLVGGVLAALAAAYCTTGSV 1680
DB 1621 PTPELYRLGAVONEVILLTHPIITKYIMACMSADLEVYVTSWTWLVGGVLAALAAAYCTTGSV 1680
QY 1681 VIVGRIILSGKPAVVDREVLYCEDEMECEASQLPYIQQGMOLAEQPKOKALGLLQATAT 1740

Db 1681 VIVGRIILSGKDAVVPDREVLYQEPDEMBECASQLPVIEQGMQLAQFQKQKALGILQITAT 1740
QY 1741 KQERAAAPVVEKWALETWFAKEMWNEFISGIQYLAGLSLTPCNPAIASIMAFASITSP 1800
Db 1741 KQERAAAPVVEKWALETWFAKEMWNEFISGIQYLAGLSLTPCNPAIASIMAFASITSP 1800
QY 1801 LITQNTILLNIGLGGVAAQIAPESAASAFVAGAGIAGAAGVSGISGLGVLDVILAGYGAGVA 1860
Db 1801 LITQNTILLNIGLGGVAAQIAPESAASAFVAGAGIAGAAGVSGISGLGVLDVILAGYGAGVA 1860
QY 1861 GALVAFKMSGEVPSPTEDLIVNLLPAILSPGALVGVVCAAILRRHVGPGEVQVWNRRI 1920
Db 1861 GALVAFKMSGEVPSPTEDLIVNLLPAILSPGALVGVVCAAILRRHVGPGEVQVWNRRI 1920
QY 1921 AFASRGNHVSPTHVPESDAAARVTOQLSLTITOLLKHLHOWINEDCSTPCSGSLRDV 1980
Db 1921 AFASRGNHVSPTHVPESDAAARVTOQLSLTITOLLKHLHOWINEDCSTPCSGSLRDV 1980
QY 1981 WDWICTVLTDFTKWLQSKLLPRLPGVPFFLSCRGYKGVWREGDGMOTTCPCGAQIAGHVK 2040
Db 1981 WDWICTVLTDFTKWLQSKLLPRLPGVPFFLSCRGYKGVWREGDGMOTTCPCGAQIAGHVK 2040
QY 2041 NGSNRIVGPRCTCNTWHTGTPPINAYTTGPTCTSPADPNYSALWRVAABEYVEVTRVGDHF 2100
Db 2041 NGSNRIVGPRCTCNTWHTGTPPINAYTTGPTCTSPADPNYSALWRVAABEYVEVTRVGDHF 2100
QY 2101 YVTGMTTNDVKPCQVPABEFTTEVDGVLHRYAPACKPLLEDVTFVGLNOLYVGSOL 2160
Db 2101 YVTGMTTNDVKPCQVPABEFTTEVDGVLHRYAPACKPLLEDVTFVGLNOLYVGSOL 2160
QY 2161 PCEPEPQVTVLTSMLTDPSSHITAKRRLARGSPPSLASSASQLSAPSLKATCTTHD 2220
Db 2161 PCEPEPQVTVLTSMLTDPSSHITAKRRLARGSPPSLASSASQLSAPSLKATCTTHD 2220
QY 2221 SPADLIHANLLWRQEMGNITVESENKVVILDSPEPLHAGDEREISVAABEILKRSK 2280
Db 2221 SPADLIHANLLWRQEMGNITVESENKVVILDSPEPLHAGDEREISVAABEILKRSK 2280
QY 2281 FPSALPIWARPDPNPPLLESKQPDVVPVHVHCPLEPTKAPITPPRRKRTVLTESNV 2340
Db 2281 FPSALPIWARPDPNPPLLESKQPDVVPVHVHCPLEPTKAPITPPRRKRTVLTESNV 2340
QY 2341 SSALAEALATKTFSSGSSAVDSGTATLPDASDDGDKGSDVSYSSMPLEGEPCDPL 2400
Db 2341 SSALAEALATKTFSSGSSAVDSGTATLPDASDDGDKGSDVSYSSMPLEGEPCDPL 2400
QY 2401 SDGSWSTVSEAEEDVVCSSNSTWTGALITPCAAEBSKLPINPLSNLLRHENMVYATT 2460
Db 2401 SDGSWSTVSEAEEDVVCSSNSTWTGALITPCAAEBSKLPINPLSNLLRHENMVYATT 2460
QY 2461 SRASLRQKVTTRDLQVLDHVRDVLKEMKAKASTVKAKLSTIEACKLTTPPHSAKSKF 2520
Db 2461 SRASLRQKVTTRDLQVLDHVRDVLKEMKAKASTVKAKLSTIEACKLTTPPHSAKSKF 2520
QY 2521 GYGAQVRLNLSRAVNHRSWEDLLEDTETPIDTTIMAKSEYFCVQPEKGRKPARLIV 2580
Db 2521 GYGAQVRLNLSRAVNHRSWEDLLEDTETPIDTTIMAKSEYFCVQPEKGRKPARLIV 2580
QY 2581 FPDGLVRVCEKMALYDVVSTLPQAVMGSSYGFQFSKQRFVEFLVNTWKSKKCPMGFSYDT 2640
Db 2581 FPDGLVRVCEKMALYDVVSTLPQAVMGSSYGFQFSKQRFVEFLVNTWKSKKCPMGFSYDT 2640
QY 2641 RCFDSTVTESDIRVEBSIYCCDLAPEARQAIISLTERLYIGGLPTNKGQNGCYRCRA 2700
Db 2641 RCFDSTVTESDIRVEBSIYCCDLAPEARQAIISLTERLYIGGLPTNKGQNGCYRCRA 2700
QY 2701 SGVLTTSCGNLTICYLKATAACRAAKLQDCTMLVNGDLVVICESAGTQEDAAALRAFT 2760
Db 2701 SGVLTTSCGNLTICYLKATAACRAAKLQDCTMLVNGDLVVICESAGTQEDAAALRAFT 2760
QY 2761 AMTRYSAAPPDGPQPEYDLELITSCSSNSVVAHDASGRVYVLTTRDPTPLAARAWETAR 2820
Db 2761 AMTRYSAAPPDGPQPEYDLELITSCSSNSVVAHDASGRVYVLTTRDPTPLAARAWETAR 2820

RESULT 7

Q99AU2 ID Q99AU2 PRELIMINARY; PRT; 3010 AA.
AC Q99AU2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus type 1b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31647;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=chimera of HCV-BK;
RA Thomson M., Nacimbani M., Gonzales S., Murlay K., Rehmann B.,
RA Liang J.;
RT "Analyses of viral sequences and virus-specific immune responses
RT during serial passage of an infectious hepatitis C virus serotype 1b
RT clone in chimpanzees."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF333324; AAK08509.1; --
DR PIR; A61196; A61196.
DR PIR; P00246; P00246.
DR PIR; P00804; P00804.
DR PIR; P0329; P0329.
DR KSP; P26663; INS3.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006118; F:electron transport; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR InterPro; IPR002518; HCV_NS2.

Db 1741 KQAEAAAPVVEKWRALFTFWAKHMWNFISGIIQYLACLSTPLGPNPAISMAFTASITSP 1800
 QY 1801 LTONTLLNKLGGWAAQAPASAFVAGAGIAGAAVSGIGLVVDILAGYGAGVA 1860
 Db 1801 LTQSTLLFNILGWWAAQAPASAFVAGAGIAGAAVSGIGLVVDILAGYGAGVA 1860
 QY 1861 GALVAFKMSGEVFSTEDLVNLLPALILSPGALVGVVCAAILRRHVSPGEGAVQWNRLLI 1920
 Db 1861 GALVAFKMSGEMPESTEDLVNLLPALILSPGALVGVVCAAILRRHVSPGEGAVQWNRLLI 1920
 QY 1921 AFASRGNEVSPTHVYVESDAAARVTOILLSLTTOLLKRLHOWINEDCSTPCSGWLRDV 1980
 Db 1921 AFASRGNEVSPTHVYVESDAAARVTOILLSLTTOLLKRLHOWINEDCSTPCSGWLRDV 1980
 QY 1981 WDMICTVLDTFKWLOSKLAPRGVPELSCQGYKGVWEGDGMOTTCPCGAQIAGHVK 2040
 Db 1981 WDMICTVLDTFKWLOSKLAPRGVPELSCQGYKGVWEGDGMOTTCPCGAQIAGHVK 2040
 QY 2041 NGSMRIIVGPRCTCNTWHTGTFPIINAYTTGCTPSPAPNYSRALNRVAABEYVEVTRVGDPH 2100
 Db 2041 NGSMRIIVGPRCTCNTWHTGTFPIINAYTTGCTPSPAPNYSRALNRVAABEYVEVTRVGDPH 2100
 QY 2101 YVTCMTIDNVKCPQVPAPEFFTEVDGRLHRYAPACKPLREDVTFOVGLNQLYVGSOL 2160
 Db 2101 YVTCMTIDNVKCPQVPAPEFFTEVDGRLHRYAPACKPLREDVTFOVGLNQLYVGSOL 2160
 QY 2161 PCPEPDVTLTSMITDPSHITATKRLRLARGSPPSLASSASQSAPSLKATCTTHD 2220
 Db 2161 PCPEPDVTLTSMITDPSHITATKRLRLARGSPPSLASSASQSAPSLKATCTTHV 2220
 QY 2221 SPDAIEANLLNRQENGNITRVESNKKVILDSFPEPLHAEDERISVAABEILKRSK 2280
 Db 2221 SPDAIEANLLNRQENGNITRVESNKKVILDSFPEPLHAEDERISVAABEILKRSK 2280
 QY 2281 FPSALPIWAPDYNPLLESKWDYDYPVPHGCPPLPTTAPIPPRKRTVVLVESNV 2340
 Db 2281 FPSALPIWAPDYNPLLESKWDYDYPVPHGCPPLPTTAPIPPRKRTVVLVESNV 2340
 QY 2341 SSALAEALATYTCSSGSSAVDSGTATAPLADSDGDKGSDVESYSSMPLEGEPPDL 2400
 Db 2341 SSALAEALATYTCSSGSSAVDSGTATAPLADSDGDKGSDVESYSSMPLEGEPPDL 2400
 QY 2401 SDGSWTSVSEAGEDVVCSSMSYTWTCALITPCAAEESKLPINPLSNLLRHNNVYATT 2460
 Db 2401 SDGSWTSVSEAGEDVVCSSMSYTWTCALITPCAAEESKLPINPLSNLLRHNNVYATT 2460
 QY 2461 SRASLRQKVTTRDLOVLDHVDVLEKMAKASTYKAKLLSTEBACKLTPPHSAKSF 2520
 Db 2461 SRASLRQKVTTRDLOVLDHVDVLEKMAKASTYKAKLLSTEBACKLTPPHSAKSF 2520
 QY 2521 GYGAKDVRLNLSRAVNHRSWEDLLEDTPIDTTIMAKSEVFCVQPEKGGKRPALIV 2580
 Db 2521 GYGAKDVRLNLSRAVNHRSWEDLLEDTPIDTTIMAKSEVFCVQPEKGGKRPALIV 2580
 QY 2581 FPDIGVRVCKMALYDVVSTLPQVWVGSSYGFQSPQQRVEFLVNTWKSKNPMGFSDYT 2640
 Db 2581 FPDIGVRVCKMALYDVVSTLPQVWVGSSYGFQSPQQRVEFLVNTWKSKNPMGFSDYT 2640
 QY 2641 RCFDSTVTEGDIIVREESIIYCCDLAPARQAIKSLTERLYTGPLTNSKQNCQYRCRA 2700
 Db 2641 RCFDSTVTEGDIIVREESIIYCCDLAPARQAIKSLTERLYTGPLTNSKQNCQYRCRA 2700
 QY 2701 SGVLTTSCGNLTLCYLKATAACRAKQDCTMLVNGDDLVIICESAGTOEDAAALAPTE 2760
 Db 2701 SGVLTTSCGNLTLCYLKATAACRAKQDCTMLVNGDDLVIICESAGTOEDAAALAPTE 2760
 QY 2761 AMTRYSAAPPDPPPEYDLELITSCSSNVSAHDASGRVYVLRDPTTPLARAWEATAR 2820
 Db 2761 AMTRYSAAPPDPPPEYDLELITSCSSNVSAHDASGRVYVLRDPTTPLARAWEATAR 2820
 QY 2821 HTPVNSWLGNIIMTAPILWARMILMTHFFSILLAQOLEKALDQCIYACYSIEPLDLPQ 2880
 Db 2821 HTPVNSWLGNIIMTAPILWARMILMTHFFSILLAQOLEKALDQCIYACYSIEPLDLPQ 2880

QY 2881 IIERLHGLSAFTLHSYSGEINRVASCLKLGVPPLRTWHRARSVRAKLLSQGGRAAATC 2940
 Db 2881 IIERLHGLSAFTLHSYSGEINRVASCLKLGVPPLRVWHRARSVRALLSQGGRAAATC 2940
 QY 2941 GYLFENWAVTKLTIPTIPAAQOLDLSCWFVAGYSGGDIYHLSRARPWFPLCULLLSV 3000
 Db 2941 GYLFENWAVTKLTIPTIPAAQOLDLSCWFVAGYSGGDIYHLSRARPWFPLCULLLSV 3000
 QY 3001 GVGIVYLLPNR 3010
 Db 3001 GVGIVYLLPNR 3010

RESULT 8
 Q9QIX5 PRELIMINARY; PRT; 3010 AA.
 ID Q9QIX5
 AC Q9QIX5;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD8-2;
 RX MEDLINE=20013325; PubMed=10544098;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.y., Miyasaka Y.,
 RA Tazawa J.i., Izumi N., Marumo F., Sato C.;
 RT "Time-related changes in full-length hepatitis C virus and hepatitis
 activity."
 RL Virology 263:244-253 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD8-2;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
 RA Sakamoto N., Fukuna T., Tazawa J., Izumi N., Marumo F., Sato C.;
 RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL: AF165060; AAD56195.1; -;
 DR PIR: A61196; A61196.
 DR PIR: PQ0246; PQ0246.
 DR PIR: PQ0254; PQ0254.
 DR PIR: PQ0804; PQ0804.
 DR PIR: PS0329; PS0329.
 DR HSSP: P26663; 1JXP.
 DR MEROPS: S29_002; -;
 DR GO: GO:0016021; C: integral to membrane; IEA.
 DR GO: GO:0019028; C: viral capsid; IEA.
 DR GO: GO:0019031; C: viral envelope; IEA.
 DR GO: GO:0005524; F: ATP binding; IEA.
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO: GO:0003723; F: RNA binding; IEA.
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F: structural molecule activity; IEA.
 DR GO: GO:0016740; F: transferase activity; IEA.
 DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P: transcription; IEA.
 DR GO: GO:0019079; P: viral genome replication; IEA.
 DR GO: GO:0019087; P: viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.

DR	InterPro: IPR002518; HCV NS2.	
DR	InterPro: IPR000745; HCV NS4b.	
DR	InterPro: IPR001490; HCV NS4b.	
DR	InterPro: IPR002868; HCV NS5a.	
DR	InterPro: IPR002166; HCV RdRp.	
DR	InterPro: IPR001650; Helicase C.	
DR	InterPro: IPR004109; Peptidase C29.	
DR	InterPro: IPR007095; RNA pol DS PS.	
DR	InterPro: IPR007094; RNA pol PS vir.	
DR	Pfam: PF01543; HCV capsid; 1.	
DR	Pfam: PF01542; HCV core; 1.	
DR	Pfam: PF01539; HCV env; 1.	
DR	Pfam: PF01560; HCV NS1; 1.	
DR	Pfam: PF01538; HCV NS2; 1.	
DR	Pfam: PF02907; HCV NS3; 1.	
DR	Pfam: PF01006; HCV NS4a; 1.	
DR	Pfam: PF01001; HCV NS4b; 1.	
DR	Pfam: PF01506; HCV NS5a; 1.	
DR	Pfam: PF020271; Helicase C; 1.	
DR	Pfam: PF00998; Viral RdRp; 1.	
DR	ProDom: PD186062; HCV NS1; 1.	
DR	SMART: SMC0487; DEXDC1; 1.	
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;	
KW	Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.	
SQ	SEQUENCE 3010 AA; 327294 MW; 8599F1E3BA6A50F56 CRC64;	
Query Match 96.1%; Score 15381; D3 12; Length 3010;		
Best Local Similarity 95.2%; Pred. No. 0;		
Matches 2866; Conservative 64; Mismatches 80; Indels 0; Gaps 0;		
QY	1 MSTNPKPQRTKNTNRPPDVKFPGGQIVGGYLLPRRGLGVATPKASERSQPRG 60	
DB	1 MSTNPKPQRTKNTNRPPDVKFPGGQIVGGYLLPRRGLGVATPKASERSQPRG 60	
QY	61 RRPIPKARPEGRNAQCPYPLVNEGLNAGWLLSPRGSRPSPGPTDPRRSNTLG 120	
DB	61 RRPIPKARPEGRNAQCPYPLVNEGLNAGWLLSPRGSRPSPGPTDPRRSNTLG 120	
QY	121 KVIDTTCGFADLMGYIPLVGAFLGGAARALAHGVRLDGVNATGNLPGCSFSIFLLA 180	
DB	121 KVIDTTCGFADLMGYIPLVGAFLGGAARALAHGVRLDGVNATGNLPGCSFSIFLLA 180	
QY	181 LLSCLTIPASAYVRNVSGIHYHTNDCSNSIYVEAADVIMHPGCVPCVQRGNSRCWV 240	
DB	181 LLSCLTIPASAYVRNVSGIHYHTNDCSNSIYVEAADVIMHPGCVPCVQRGNSRCWV 240	
QY	241 ALPTTAAARNASVPTTIRRHVDLLVGTAAFCSAMVYVDLCGSIFLVLSQFTFSPRRHET 300	
DB	241 ALPTTAAARNISVPTTIRRHVDLLVGTAAFCSAMVYVDLCGSFVLVSQFTFSPRRHET 300	
QY	301 VQDCNSIYPGVSHGHMAMWMMNSPTTALVVSQLLRPQAVDMVAGAHGVLAGLA 360	
DB	301 VQDCNSIYPGVSHGHMAMWMMNSPTAALVVSQLLRIPQAVDMVAGAHGVLAGLA 360	
QY	361 YYSVMGNWAKVLIVALLFAGVDGTHTRVAGHTTSGFTSLPSSGASOKIOLVNTNGSW 420	
DB	361 YYSVMGNWAKVLIVMLLFAGVDGTHTRVGAQHSANTLSFTVGPQKIQVNTNGSW 420	
QY	421 HINRTALNCNDSLQTFGAALFYAHKFNSSCGCPERVASCRPIDWFAGCGWGPITYTKPNSS 480	
DB	421 HINRTALNCNDSLQTFGAALFYTHRNASGCPERVASCRPIDTFDGGWGPITYABNSP 480	
QY	481 DQRPYCHWYAPRCGVVPASQGVGYCTFTPSVVGTTDRSGVPTVSGENETDVMLLN 540	
DB	481 DQRPYCHWYAPRCGIEPASKVGVGYCTFTPSVVGTTDRDFGVPYTNWGENETDVMLLN 540	
QY	541 NTRPPQGNWFGCTWMNSTGTCTCGGPPCNIGVGNRTLCPTDCSKPKHPEAFTYTKCGSG 600	
DB	541 NTRPPQGNWFGCTWMNSTGTCTCGGPPCNIGAGNNTLCPTDCFKKHPEAFTYTKCGSG 600	
QY	601 PWLTPCLVDYPYRLWHYPCTLNFSPKVRMYVGGVEHRLNAACNWTGRCNLEDNRDS 660	
DB	601 PWLTPCLVDYPYRLWHYPCTVNFIPKVRMYVGGVEHRLSAACNWTGRCNLEDNRDS 660	

QY	661 ELSPLLSTTEWQIILPCAFITTLPALSTGLIHLHONIVDQVLYGVSAFVSFAIKWEYIL 720	
DB	661 ELSPLLSTTEWQIILPCSFITTLPALSTGLIHLHONIVDQVLYGVSAFVSFAIKWEYIL 720	
QY	721 LFLFLLLADARVCACILWMLLTAQAEAALENLVLNAAVAGAHGILSELVFFCAAWYIKG 780	
DB	721 LFLFLLLADARVCACILWMLLTAQAEAALENLVLNAAVAGAHGILSELVFFCAAWYIKG 780	
QY	781 RLAPGAAYAFYGVWP:LLLLLLALPPRAVALDREMAASCGGAVLVGLVFLTSLPYKVFIL 840	
DB	781 RLVPGAAYALYGVWPELLLLLLALPPRAYANDREMAASCGGAVFVGLAFLLSPYKVFILA 840	
QY	841 RLINWLQYFITRAEAHMOWPPLNVRGGRDAIILLTCVHPHLLFDITKLLALLIGPLM 900	
DB	841 RLINWLQYFITRAEAHLVWVPEPLNVRGGRDAIILLMCVHPHLLFDITKLLALLIGPLM 900	
QY	901 VLQAGITRVYFVRAQGLIRACMLVKRVAGGHYVQVFMKLGALTGTYYVNHLPBLDKWA 960	
DB	901 VLQAGITRVYFVRAHGLIRACMLVKRVAGGHYVQVMAFKLAALTGTYYVNDHLAQHWA 960	
QY	961 HAGRLAVAVEPVVFSAMETKVIITWADTAACGDIIILGLPVSARRGKEIFLGPAUSLEG 1020	
DB	961 HGLRDLAVAVPVVFSKDVETKIIITWADTAACGDIIILGLPVSARRGRELILGPADSLEG 1020	
QY	1021 QGMRLLAPITAYSQQTRGVGLGCIITSLTGRDXQVGEVQVWVSTATQSFATCINGVCWT 1080	
DB	1021 QGMRLLAPITAYSQQTRGLLGCIITSLTGRDXQVGEVQVWVSTATQSFATCINGVCWT 1080	
QY	1081 VVHGAGSKTLAPKGPITOMYTNVDLDLVGWCAAPGASRMTFCSCGSDLYLVTHADVI 1140	
DB	1081 VVHGAGSKTLAPKGPITOMYTNVDQLVWCAAPGASRMTFCSCGSDLYLVTHADVI 1140	
QY	1141 PVRRRDSRGSLLSPRPVSYLKSGSGPLLCPSGHVGVVFRAAVCTRGVAKAVDIPVPS 1200	
DB	1141 PVRRRDSRGSLLSPRPVSYLKSGSGPLLCPSGHAVGIFRAAVCTRGVAKAVDIPVPS 1200	
QY	1201 MBTTXRSPVFTDNSTPPAVPQTFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLVLPNSVAA 1260	
DB	1201 MBTTXRSPVFTDNSPPAVPQTFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLVLPNSVAA 1260	
QY	1261 TLGFGAYMKAHIGIDPNIRGTITGSGIITYSYGKFLADCGGCGGAYDIIICDECHS 1320	
DB	1261 TLGFGAYMKAHIGIDPNIRGTITGSGIITYSYGKFLADCGGCGGAYDIIICDECHS 1320	
QY	1321 TDSTTILGTGTVLDQAETAAGARLVLAATATPPGSVTVPEPNTIEEIGLSNNGSIPFYGKAI 1380	
DB	1321 TDSTTILGTGTVLDQAETAAGARLVLAATATPPGSVTVPEPNTIEEVALSNTGIPFYGKAI 1380	
QY	1381 PIRAIKGGRHILFCHSKKKCDLAAKLTGLGNNAVAYRGLDVSVIPIPGDVVVVATDAL 1440	
DB	1381 PIEVIKGGRHILFCHSKKKCDLAAKLSALGNNAVAYRGLDVSVIPTSGDVVVVATDAL 1440	
QY	1441 MTGFTGDPDSVIDCMTCVTQTVDFSLDPTFTLETTTPQDAVSRSORCGRGSGGIYR 1500	
DB	1441 MTGYTGDFDSVIDCMTCVTQTVDFSLDPTFTLETTTPQDAVSRQRGRGSGGIYR 1500	
QY	1501 FVTPGERPSGMPDSSVLCECYDAGCAWYELTAPETSVRLRAYINTPGLPVCDHLEFWS 1560	
DB	1501 FVTPGERPSGMPDSSVLCECYDAGCAWYELTAPETSVRLRAYINTPGLPVCDHLEFWS 1560	
QY	1561 VFTGLTHIDAHFLSQTKQAGDNFPYLVAQVAVCARAAPPSPSQMCKLIRLPTLHG 1620	
DB	1561 VFTGLTHIDAHFLSQTKQAGDNFPYLVAQVAVCARAAPPSPSQMCKLIRLPTLHG 1620	
QY	1621 EPIPLRYLCAVONEVILTHPIYKIMACMSADLEVVTSTWLVGGVLAALAAAYCIJTSGV 1680	
DB	1621 EPIPLRYLCAVONEVILTHPIYKIMACMSADLEVVTSTWLVGGVLAALAAAYCIJTSGV 1680	
QY	1681 VIVGRILLSGKPAVDPREVLYQEFDEMEECASQLPYIEQGMQLAEQPKQKALGLQTAT 1740	
DB	1681 VIVGRILLSGKPAVDPREVLYQEFDEMEECASHLPYIEQGMQLAEQPKQKALGLQTAT 1740	

1741 KOAEAAAPVVEKSWRALETFWAKHMWNFIISGQYLAGLSTLPCNPALIASLMAFTSITSP 1800
Db KOAEAAAPVVEKSWRALETFWAKHMWNFIISGQYLAGLSTLPCNPALIASLMAFTSITSP 1800
1801 LTTQNTLLNIIIGWVAAQIAPPASAFVAGIAGAAGVSGTGLKVLVDIILAGYAGVA 1860
Db LTTQNTLLNIIIGWVAAQIAPPASAFVAGIAGAAGVSGTGLKVLVDIILAGYAGVA 1860
1861 GALVAFKWSGEVPESTEDLVNLLPAILSPGALVGVVCAAILPRHYGPGRGAVQVMMNRLL 1920
Db GALVAFKWSGEVPESTEDLVNLLPAILSPGALVGVVCAAILPRHYGPGRGAVQVMMNRLL 1920
1921 AFASRGNHVSPTHYVPESDAAARVTOILSLTTTQLLKRHLHWINBDCSTPCSGMLRDV 1980
Db AFASRGNHVSPTHYVPESDAAARVTOILSLTTTQLLKRHLHWINBDCSTPCSGMLRDV 1980
1981 WDMICVLTDFKTLWLSKLLPLRPGVFFLSQGRYKGVWRGDIYQTTCPGQAIAGHYK 2040
Db WDMICVLTDFKTLWLSKLLPLRPGVFFLSQGRYKGVWRGDIYQTTCPGQAIAGHYK 2040
2041 NGSMRIVGRTCSNTWHGTFPPINAYTTGCTPSPAPENYSRALWVAABEYVEVTRVGDHF 2100
Db NGSMRIVGRTCSNTWHGTFPPINAYTTGCTPSPAPENYSRALWVAABEYVEVTRVGDHF 2100
2101 YVTGMTTDNVKCPQVAPBEPFTEVDGVRHLRVAPACKPILLREVTFFQVGLNQLVGSOL 2160
Db YVTGMTTDNVKCPQVAPBEPFTEVDGVRHLRVAPACKPILLREVTFFQVGLNQLVGSOL 2160
2161 PCEPEPDVTLTSMITDPSHITAEAKRLARGSPPLSSASSASQLSAPSKATCTTHD 2220
Db PCEPEPDVTLTSMITDPSHITAEAKRLARGSPPLSSASSASQLSAPSKATCTTHD 2220
2221 SPADLIEANLLWRQEMGNITRVESKENVILDSPEPLHAEDEDEISVAAAILRKS 2280
Db SPADLIEANLLWRQEMGNITRVESKENVILDSPEPLHAEDEDEISVAAAILRKS 2280
2281 FPAALIEARPDYNNPILLESKWDPDVYPPVHGCPLPPTKAPIPPRKRRTVVLTESNV 2340
Db FPAALIEARPDYNNPILLESKWDPDVYPPVHGCPLPPTKAPIPPRKRRTVVLTESNV 2340
2341 SSALAEATKTFGSSGSSAVDSGTATAPPDQASDGDGSDVESYSMPLEGEPPDL 2400
Db SSALAEATKTFGSSGSSAVDSGTATAPPDQASDGDGSDVESYSMPLEGEPPDL 2400
2401 SDGSWSTVSEAEEDVVCSSMSYTWTCALITPCAAEESEKLPINPLNSILRHHNVYATT 2460
Db SDGSWSTVSEAEEDVVCSSMSYTWTCALITPCAAEESEKLPINPLNSILRHHNVYATT 2460
2461 SRASLRQKVTFRLOVLDHVRDVLKEMKAKASTVKAKLLSTEEACKLTPPHSARSKF 2520
Db SRASLRQKVTFRLOVLDHVRDVLKEMKAKASTVKAKLLSTEEACKLTPPHSARSKF 2520
2521 GYKAKVRNLSRAVNHIRSVWEDLLEDTETPIDDITINAKSEVFCVPEKGGKPARLIV 2580
Db GYKAKVRNLSRAVNHIRSVWEDLLEDTETPIDDITINAKSEVFCVPEKGGKPARLIV 2580
2581 FPDGLVRVCBKALYDVVSTLPQAVMGSSYGFQSPKQRFVFLVNTWKSCKCPMGFSYDT 2640
Db FPDGLVRVCBKALYDVVSTLPQAVMGSSYGFQSPKQRFVFLVNTWKSCKCPMGFSYDT 2640
2641 RCFDSTVTESDIRVESIYQCCDLAPBARQAIRSLTERLITGGLTNSKQNCYRRCRA 2700
Db RCFDSTVTESDIRVESIYQCCDLAPBARQAIRSLTERLITGGLTNSKQNCYRRCRA 2700
2701 SGVLTSCGNLTLCYLKATACRAAKLODCFMVNGDVLVVICESAGTOEDARALAPTE 2760
Db SGVLTSCGNLTLCYLKATACRAAKLODCFMVNGDVLVVICESAGTOEDARALAPTE 2760
2761 AMTRYSAPPDGPPEYDLELITSCSNVSVVAHDASGKRVYLTFRDPTTPLARAAWETAR 2820
Db AMTRYSAPPDGPPEYDLELITSCSNVSVVAHDASGKRVYLTFRDPTTPLARAAWETAR 2820
2821 HTPNSWLGNIIMVAPTLWARMILMTHFFSILLAQBLEKALDCQIYGACYSIEPLDLPQ 2880

Db 2821 HTPNSWLGNIIMVAPTLWARMILMTHFFSILLAQBLEKALDCQIYGACYSIEPLDLPQ 2880
Qy 2881 IIERLHGHSAPFTLHSGVSPGEINRVASCLRKGLGVPPLRTWRHRARSVRAKLLSQGGAATC 2940
Db 2881 IIERLHGHSAPFTLHSGVSPGEINRVASCLRKGLGVPPLRTWRHRARSVRAKLLSQGGAATC 2940
Qy 2941 GSYLFNVAVRTKLKLTPIPAASQDLGSFVAGYSGGDIYHLSLSRPRWPFCLLLLSV 3000
Db 2941 GSYLFNVAVRTKLKLTPIPAASQDLGSFVAGYSGGDIYHLSLSRPRWPFCLLLLSV 3000
Qy 3001 GVGIVLLPNR 3010
Db 3001 GVGIVLLPNR 3010
RESULT 9
Q9J3H7 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD15;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
progression.";
RL Submitted (NCV-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF207756; AAF65946.1; -;
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0329; PS0329.
DR HSP; P26663; IJXP.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR003345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002869; HCV NS5a.
DR InterPro; IPR002166; HCV_RdRp.

Db 1801 LTTQSTLLFNILGGWAAQAPPSAASAFVAGAGAAVGSIGLCKVLVDILLAGYGAV 1860
QY 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVCPGEGAVQWNRLL 1920
Db 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVCPGEGAVQWNRLL 1920
QY 1921 AFASRGNHVSPTHYVPESDAARAVTQILGSLTITQLLKHQWINEDECSTPCSGSWLRDV 1980
Db 1921 AFASRGNHVSPTHYVPESDAARAVTQILGSLTITQLLKHQWINEDECSTPCSGSWLRDV 1980
QY 1981 KDWICTVLDFKTLWLSKXLLPRLPGVFFLSCQGYKGVWRGDMGIMOTPCGQAIAGHVK 2040
Db 1981 KDWICTVLDFKTLWLSKXLLPRLPGVFFLSCQGYKGVWRGDMGIMOTPCGQAIAGHVK 2040
QY 2041 NGSMTVGRPTCSNTWHGTFPPINAYTGPCTPAPNYSALWVAEEVEVTRVGDHF 2100
Db 2041 NGSMTVGRPTCSNTWHGTFPPINAYTGPCTPAPNYSALWVAEEVEVTRVGDHF 2100
QY 2101 VYTGMTTDNVKPCQVPAPEFFTEVDGVRHLRYAPACKPFLLEDVTFQVGLNQLVGSOL 2160
Db 2101 VYTGMTTDNVKPCQVPAPEFFTEVDGVRHLRYAPACKPFLLEDVTFQVGLNQLVGSOL 2160
QY 2161 PCBPEDVWVLTSMITDPSHITATKRRILARGSPPELASSASQLSAPSLKATCTTHHD 2220
Db 2161 PCBPEDVWVLTSMITDPSHITATKRRILARGSPPELASSASQLSAPSLKATCTTHHD 2220
QY 2221 SPADLIEANLWRQBMGNITRVESNKKVILDSFELHABGDEREISVAAILRKSX 2280
Db 2221 SPADLIEANLWRQBMGNITRVESNKKVILDSFELHABGDEREISVAAILRKSX 2280
QY 2281 FPSALPIWAPEDYNPPLLESKWDPPYVPPVHVGCPLPPTXAPIPPPRRKRTVVLTESNV 2340
Db 2281 FPSALPIWAPEDYNPPLLESKWDPPYVPPVHVGCPLPPTXAPIPPPRRKRTVVLTESNV 2340
QY 2341 SSALAEIATKTFGSSGSAVDGSTATALDPLADDDGDKGSDVSYSNMPLGEPGDPDL 2400
Db 2341 SSALAEIATKTFGSSGSAVDGSTATALDPLADDDGDKGSDVSYSNMPLGEPGDPDL 2400
QY 2401 SDGWSVTSSEAGEDVVCSSSVTWTCALITPCAAEESKLPINPLNSLLRHNNVYATT 2460
Db 2401 SDGWSVTSSEAGEDVVCSSSVTWTCALITPCAAEESKLPINPLNSLLRHNNVYATT 2460
QY 2461 SRASLRQKQKVTEDRLQVLDCHYSDVLKEMKAKASTVYKALLSIEEACKLTPPHSAKSF 2520
Db 2461 SRASLRQKQKVTEDRLQVLDCHYSDVLKEMKAKASTVYKALLSIEEACKLTPPHSAKSF 2520
QY 2521 GYGAQVNRNLSRAVNHIRSWEDLLEDTEPTDITIMAKSEVFCVQPEKGRKPARLIV 2580
Db 2521 GYGAQVNRNLSRAVNHIRSWEDLLEDTEPTDITIMAKSEVFCVQPEKGRKPARLIV 2580
QY 2581 FPDGLGVCEKMAIYDVVSTLPQAVMGSSYGFQYSPKORVEFLVNTWKSCKPMGFSYDT 2640
Db 2581 FPDGLGVCEKMAIYDVVSTLPQAVMGSSYGFQYSPKORVEFLVNTWKSCKPMGFSYDT 2640
QY 2641 RCFDSTVTSIEDIRVEESIYCCDLAPAEARQAIRSLTERLYIGGFLTNKSGONCGYRCRA 2700
Db 2641 RCFDSTVTSIEDIRVEESIYCCDLAPAEARQAIRSLTERLYIGGFLTNKSGONCGYRCRA 2700
QY 2701 SGVLTTCGNTLTICYLATAACRAKQLODCTMLVNGDDLWVICSAGTQEDAAALRAFTE 2760
Db 2701 SGVLTTCGNTLTICYLATAACRAKQLODCTMLVNGDDLWVICSAGTQEDAAALRAFTE 2760
QY 2761 AMTRYAPPQDPPEVDLELITSCSSNVSVAHASGKRVYLTREDPTTPLARAWEATAR 2820
Db 2761 AMTRYAPPQDPPEVDLELITSCSSNVSVAHASGKRVYLTREDPTTPLARAWEATAR 2820
QY 2821 HTPINSMIGNIIMYAPTILWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQ 2880
Db 2821 HTPINSMIGNIIMYAPTILWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQ 2880
QY 2881 IIEELHGLSAPLTHSYSPGENRVASCLRLKGVPLRTWRHARSVRAKLLSQGGRATC 2940
Db 2881 IIEELHGLSAPLTHSYSPGENRVASCLRLKGVPLRTWRHARSVRAKLLSQGGRATC 2940

QY 2941 GRYLFENWAVRTKLTPIPAASQLDSGWFVAGYSGGDIYHLSRARPWFPLCLLLLSV 3000
Db 2941 GRYLFENWAVRTKLTPIPAASQLDSGWFVAGYSGGDIYHLSRARPWFPLCLLLLSV 3000
QY 3001 GVGIIYLLPNR 3010
Db 3001 GVGIIYLLPNR 3010
RESULT 10
Q9J3H5
ID Q9J3H5 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H5;
DT 01-OCT-2000 (trEMBLrel. 15, Created)
DT 01-OCT-2000 (trEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (trEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
SEQUENCE FROM N.A.
RP
RC STRAIN=MD17;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DDRJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (3Y SIMILARITY).
CC EMBL: AF207758; AAF65948.1; -
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PQ0254; PQ0254.
DR PIR; PS0329; PS0329.
DR HSP; P27958; LHE1.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR000345; Cyt heme_Bs.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR004109; Peptidase C29.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.

DR pfam; PF01542; HCV core; 1.
 DR pfam; PF01539; HCV env; 1.
 DR pfam; PF01560; HCV NS1; 1.
 DR pfam; PF01538; HCV NS2; 1.
 DR pfam; PF02907; HCV NS3; 1.
 DR pfam; PF01006; HCV NS4a; 1.
 DR pfam; PF01001; HCV NS4b; 1.
 DR pfam; PF01506; HCV NS5a; 1.
 DR pfam; PF00271; helicase C; 1.
 DR pfam; PF00398; viral RdRp; 1.
 DR ProDom; PD186082; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00130; CYTOCHROME C; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Glycoprotein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326801 MW; 95EE3D1B93B7AA4B CRC64;

Query Match 95.9%; Score 15358; D3 12; Length 3010;
 Best Local Similarity 95.5%; Pred. No. 0;
 Matches 2874; Conservative 57; Mismatches 78; Indels 2; Gaps 2;

QY 1 MSTNPEQKTKENTNRRPDQVKFPGGQIVGVYLLPRGRLGVATRKASERSQPRG 60
 DB 1 MSTNPEQKTKENTNRRPDQVKFPGGQIVGVYLLPRGRLGVATRKASERSQPRG 60

QY 61 RQPIEKARRPEGRAWAQPGYPWLYGNEGLWAGWLLSPRGSRPSWGPTDPRRRSRLG 120
 DB 61 RQPIEKARQPEGRAWAQPGYPWLYGNEGMWAGWLLSPRGSRPSWGPTDPRRRSRLG 120

QY 121 KVIDTLTQCFADLMGYPIPLVGAELGGAARALAEGRVLEEDGVNYATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTQCFADLMGYPIPLVGAELGGAARALAEGRVLEEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLPIPASAYVRNVSGLHYHNTDCNSNSIYVEADVIMHTPCVPCVQVQEGNSRCVW 240
 DB 181 LLSCLVPSAYEVRNVSGAHVNTDCNSNSIYVETADMIMHTPCVPCVQVQEGNSRCWA 240

QY 241 ALPTFLAARNASVPTTIRRHVDLLVGTAFCSAMVVGDLGSGIFLVSOFTFSPRRHET 300
 DB 241 ALPTFLAARNASVPTTIRRHVDLLVGAAFCSAMVVGDLGSGVFLVLSQFTFSPRRHET 300

QY 301 VQDCNGSIYPGHVSGHMAWDMNNHSPPTALVVSOLLRIPOAVDMVGAHGVLAGLA 360
 DB 301 VQDCNGSIYPGHVSGHMAWDMNNHSPPTALVVSOLLRIPOAVDMVGAHGVLAGLA 360

QY 361 YYSMWGNWAKLVALLFAGVQGE-THTCRVAGHTSGFTSLFSSCASOKIOLVNTNGS 419
 DB 361 YYSMIGNWAKLVLLFAGVQGTGLVSGVSGHTTSGLTSLFRVGCASQRIOLVNTNGS 420

QY 420 WHINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERYASCRPIDWPAQGWGPTITYTKPNS 479
 DB 421 WHINRTALNCNDSLOTGFFAALFYTNKFNSSGCSERLNSCRPIENTTQGWGPTITYDEGG 480

QY 480 SDORPCWHYAPPCGVWASQVCGPVYCTPSPVWVWTDGSGVPTYSGENETDMLL 539
 DB 481 SDORPCWHYAPQCGVIVPASEVCGPVYCTPSPVWVWTDGSGVPTYSGENETDMLL 540

QY 540 NNTRPPQGNWFGCTMNSGTCTKTCGGPPCNIGVGNRTLICTDCFRKHPEATYTKCGS 599
 DB 541 NNTRPPQGNWFGCTMNSGTCTKTCGGPPCNIGVGNR-LVCFDTCFRKHPEATYTKCGS 599

QY 600 GPWLTPRCLVDYPIRLWHYPCTTLNFSIFKVRMYVGGVEHRLNAACNWTGERCNLEDRDR 659
 DB 600 GPWLTPRCLVDYPIRLWHYFCTVNTYVFKVRMYVGGVEHRLSAACKWTRGERCNLEDRDR 659

QY 660 SELSPILLSTTEWQILLPCATITLIPALSTGLHLHQNIVDVQYLYGVGSAPVSPAIKWEYI 719
 DB 660 SELSPILLSTTEWQILLPCSPITLIPALSTGLHLHQNIVDVQYLYGVGSAPVSPAIKWEYI 719

QY 720 LLLFLLLLADARVCACLLWMLLFAQEAALENLVVLNNAASVAGAHGILSLVFFCAAWYIK 779
 DB 720 LLLFLLLLADARVCACLLWMLLKPQEAALENLVVLNNAASVAGAHGILSLVFFCAAWYIK 779

QY 780 GRLAPGAAYAFVGVWPLLLLLLPPRAYALDREMAASCGGAVILVGLVFLTLSPYKVF 839
 DB 780 GRLAPGAAYAFVGVWPLLLLLLPPRAYALDREMAASCGGAVFVGLVFLTLSPYKVF 839

QY 840 TRLIWLQYFTTRAHAHQVWVPPINVGGSDAILLTCAVHPELIDITKLLAILGL 899
 DB 840 TRLIWLQYFTTRAHAHQVWVPPINVGGSDAILLTCAVHPELIDITKLLAILGL 899

QY 900 MYLQAGITRVYFVFAOGLIRACMLVRKVAGHYVQVMFKLGALTGVVYVNHITPLRDW 959
 DB 900 MYLQAGITRVYFVFAOGLIRACMLVRKVAGHYVQVMFKLGALTGVVYVNHITPLRDW 959

QY 960 AHAGLRDLAVAVEPVVFSAMETKVITWGAATAACGDIILGLPVSARRKEIFLGPADSLE 1019
 DB 960 AHAGLRDLAVAVEPVVFSAMETKVITWGAATAACGDIILGLPVSARRKEIFLGPADSLE 1019

QY 1020 GQGWELLAPITAYSQOTRGLLCIITSLTGDDKNQVGEVQVVSATQSFATCTINGVCM 1079
 DB 1020 GQGWELLAPITAYSQOTRGLLCIITSLTGDDKNQVGEVQVVSATQSFATCTINGVCM 1079

QY 1080 TYTHGAGSKTLGAPKGPITQMTNVVDLGVWQAPPGARSMTPCSGSSDLYLVTRHADV 1139
 DB 1080 TYTHGAGSKTLGAPKGPITQMTNVVDLGVWQAPPGARSMTPCSGSSDLYLVTRHADV 1139

QY 1140 IPVRRGDSRGSLLSPRPVSYLKGSSGGFLLCPSGHVVGVFRAAYCTRGVAKAVDFIPVE 1199
 DB 1140 IPVRRGDSRGSLLSPRPVSYLKGSSGGFLLCPSGHVVGVFRAAYCTRGVAKAVDFIPVE 1199

QY 1200 SMETMRSVPFTDNTTPPAVQTFQVAHLHAPTSGSKSTKVPAAVAAQYKVLNPSVA 1259
 DB 1200 SMETMRSVPFTDNTTPPAVQTFQVAHLHAPTSGSKSTKVPAAVAAQYKVLNPSVA 1259

QY 1260 ATLGFGAYMSKAEGDDPNIRGTGRTITTTGGSTYTYGKFLADGCGSGAVDIIICDECH 1319
 DB 1260 ATLGFGAYMSKAEGDDPNIRGTGRTITTTGGSTYTYGKFLADGCGSGAVDIIICDECH 1319

QY 1320 STDSTTILGIGTVLQDAETAGARLVLTATPPGSVTVPHPNIEIGLSNNGEIPFYGKA 1379
 DB 1320 STDSTTILGIGTVLQDAETAGARLVLTATPPGSVTVPHPNIEIGLSNNGEIPFYGKA 1379

QY 1380 IPIEAIKGRHLIPCHSKKCDLAAKLTGLNNAVAYRGLDVSVIPDGVVVVATDA 1439
 DB 1380 IPIEVIKGRHLIPCHSKKCDLAAKLTGLNNAVAYRGLDVSVIPDGVVVVATDA 1439

QY 1440 LMTGTGDPDSVIDCNTCVTQVDSLDPTFTIETTTVPQDAVSRQRGRGSGTY 1499
 DB 1440 LMTGTGDPDSVIDCNTCVTQVDSLDPTFTIETTTVPQDAVSRQRGRGSGTY 1499

QY 1500 RFTVTPGERPSGVFDSVILCECYDAGCAYELTPAETSURLRAYLNTPLGVQODHLERWE 1559
 DB 1500 RFTVTPGERPSGVFDSVILCECYDAGCAYELTPAETSURLRAYLNTPLGVQODHLERWE 1559

QY 1560 SVPTGCTHIDAHFLQTKOAGNPPYLVAQATVCARAQAPPSWDQWKCLIRLKPILH 1619
 DB 1560 SVPTGCTHIDAHFLQTKOAGNPPYLVAQATVCARAQAPPSWDQWKCLIRLKPILH 1619

QY 1620 GPTPLLYRLGAVQNEVLTHTPTIKVIMCMSADLEVTSTWLVGVGLAALAAAYCLTTGS 1679
 DB 1620 GPTPLLYRLGAVQNEVLTHTPTIKVIMCMSADLEVTSTWLVGVGLAALAAAYCLTTGS 1679

QY 1680 VVIVGRILISGKPAVPPDEVLYQSFDEMEBCASQPIYIEQGMQLAEQFKQALGLQTA 1739
 DB 1680 VVIVGRILISGKPAVPPDEVLYQSFDEMEBCASQPIYIEQGMQLAEQFKQALGLQTA 1739

QY 1740 TYQAAAPAVESKWRALFTFWAKEMWNPISGIVLAGLSTLPGNPALASLMAFTASITS 1799
 DB 1740 TYQAAAPAVESKWRALFTFWAKEMWNPISGIVLAGLSTLPGNPALASLMAFTASITS 1799

QY 1800 PLTTQNTLLFNILGQWVAQAPPSAASAFVAGIAGAAGVSGIGKVLVDILAGYGAGV 1859
 DB 1800 PLTTQNTLLFNILGQWVAQAPPSAASAFVAGIAGAAGVSGIGKVLVDILAGYGAGV 1859

QY 1860 AGALVAFKMGSEVPSSTEDLVNLLFAILSPGALVGVVCAALIRHVGPGEGAVOMARL 1919
Db 1860 AGALVAFKMGSEVPSSTEDLVNLLFAILSPGALVGVVCAALIRHVGPGEGAVOMARL 1919
QY 1920 IAFASRGNHVSPHVVPSDAAARVQIILSSITITOLLKRLHOWINEDCSIFCSGWLDR 1979
Db 1920 IAFASRGNHVSPHVVPSDAAARVQIILSSITITOLLKRLHOWINEDCSIFCSGWLDR 1979
QY 1980 VMDWICTVLTDFKWLQSKLPLRPGVPLFLSCORGYKGVWRGDTQMOTTCPGAQIAGHV 2039
Db 1980 VMDWICTVLTDFKWLQSKLPLRPGVPLFLSCORGYKGVWRGDTQMOTTCPGAQIAGHV 2039
QY 2040 KNGSMRIIVGPRCTSNWHTGTPINAYTTGCPPTSPAPNYSRALMRVAEEVVEVTRVGF 2099
Db 2040 KNGSMRIIVGPRCTSNWHTGTPINAYTTGCPPTSPAPNYSRALMRVAEEVVEVTRVGF 2099
QY 2100 FYVTGMTIDNVKPCQVPAPEFFTEVDGVRHRYAPACKPRLREDVTFQVGLNVLVGSQ 2159
Db 2100 FYVTGMTIDNVKPCQVPAPEFFTEVDGVRHRYAPACKPRLREDVTFQVGLNVLVGSQ 2159
QY 2160 LPCPEPDVTVLTMIDPDSHITAEAKRLARGSPPSLASSASQLSAPSLKATCTTHH 2219
Db 2160 LPCPEPDVTVLTMIDPDSHITAEAKRLARGSPPSLASSASQLSAPSLKATCTTHH 2219
QY 2220 DSPADLIEANILKQEWGCMITRVESENKVILDSFPLHABGDERISVAABILRKSR 2279
Db 2220 DSPADLIEANILKQEWGCMITRVESENKVILDSFPLHABGDERISVAABILRKSR 2279
QY 2280 KFPSPALPIWAPDYN2PULLESWKDDYVPPVHVGCELPPTKAPPIPPRRKRTVVLDES 2339
Db 2280 KFPSPALPIWAPDYN2PULLESWKDDYVPPVHVGCELPPTKAPPIPPRRKRTVVLDES 2339
QY 2340 VSSALAEIATKTFSSGSSAVDSGTATLPDLASDDGDKGSDVESYSSMPLEGEPPD 2399
Db 2340 VSSALAEIATKTFSSGSSAVDSGTATLPDLASDDGDKGSDVESYSSMPLEGEPPD 2399
QY 2400 LSDGSWSVSEAEEDVCCSWSYVTWTCALITPCAAESKLPINPLNSLIRHNMVAT 2459
Db 2400 LSDGSWSVSEAEEDVCCSWSYVTWTCALITPCAAESKLPINPLNSLIRHNMVAT 2459
QY 2460 TSRSASLRQKQVTFDRLOVLDHVDVLEKEMKAKASTVKAKLLSTIEZACKLT2PHSAKSK 2519
Db 2460 TSRSASLRQKQVTFDRLOVLDHVDVLEKEMKAKASTVKAKLLSTIEZACKLT2PHSAKSK 2519
QY 2520 FGYGAKDVRLSSRAVNHIRSWEDLLEDTETPTITIMAXSEVFCVQPERGGRKPARLI 2579
Db 2520 FGYGAKDVRLSSRAVNHIRSWEDLLEDTETPTITIMAXSEVFCVQPERGGRKPARLI 2579
QY 2580 VFPLGVRVCEKALYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTWKSCKCPMGFSYD 2639
Db 2580 VFPLGVRVCEKALYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTWKSCKCPMGFSYD 2639
QY 2640 TRCFDSTVTESDIRVESIYQCCDLAEPARQAIKSLTERLYIGGLPLTNSKQNGCYRRCR 2699
Db 2640 TRCFDSTVTESDIRVESIYQCCDLAEPARQAIKSLTERLYIGGLPLTNSKQNGCYRRCR 2699
QY 2700 ASGVLTITSCGNLTICYIKATAACAAKLQDCTMLVNGDDLVIICESAGTQEDAAALRAFT 2759
Db 2700 ASGVLTITSCGNLTICYIKATAACAAKLQDCTMLVNGDDLVIICESAGTQEDAAALRAFT 2759
QY 2760 EAMTRYSAAPPDPPQPEVDLELITSCSNVSVVAHDSKRYVYLTRPTTFLAANAETA 2819
Db 2760 EAMTRYSAAPPDPPQPEVDLELITSCSNVSVVAHDSKRYVYLTRPTTFLAANAETA 2819
QY 2820 RHTPINSLWLGNIIMYAPTILWARMILMTHFFSILLAQEOLKALDCQIYGACYSIEPLDLP 2879
Db 2820 RHTPINSLWLGNIIMYAPTILWARMILMTHFFSILLAQEOLKALDCQIYGACYSIEPLDLP 2879
QY 2880 QIIRLHGLSAPTHHSYSPGNEINVASCLRLKGVPLRTWEHRARSVRKLLSOGGPAAT 2939
Db 2880 QIIRLHGLSAPTHHSYSPGNEINVASCLRLKGVPLRTWEHRARSVRKLLSOGGPAAT 2939
QY 2940 CGRYLFWAVRTKLTPIPAASQLDLSGWFWAGYSGGDIVHLSLRARPRWFPLCLLALLS 2999

Db 2940 CGRYLFWAVRTKLTPIPAASQLDLSGWFWAGYSGGDIVHLSLRARPRWFPLCLLALLS 2999
QY 3000 VGVGIYLLPNR 3010
Db 3000 VGVGIYLLPNR 3010
RESULT 11
Q9QIX6
ID Q9QIX6 PRELIMINARY; PRT: 3010 AA.
AC Q9QIX6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
[1]
RN STRAIN=MD8-1;
RP SEQUENCE FROM N.A.
RC STRAIN=MD8-1;
RA MEDLINE=20013325; PubMed=10544098;
RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
RA Tazawa J.I., Izumi N., Marumo F., Sato C.;
RA "Time-related changes in full-length hepatitis C virus and hepatitis
RT activity";
RT virology 263:244-253 (1999).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MD8-1;
RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -! SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF165059; AAD56194.1; -;
DR PIR: A61196; A61196.
DR PIR: PQ0246; PQ0246.
DR PIR: PQ0254; PQ0254.
DR PIR: PQ0804; PQ0804.
DR PIR: PS0329; PS0329.
DR HSSP: P26663; LUXP.
DR MEROPS: S29.002; -;
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0019028; C:viral capsid; IEA.
DR GO: GO:0019031; C:viral envelope; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003668; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR GO: GO:0019079; P:viral genome replication; IEA.
DR GO: GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR001410; DEAD_-.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR000745; HCV NS4a.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRp.
DR InterPro: IPR001650; Helicase_C.

```

DR InterPro: IPR004109; Peptidase C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSvir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase C; 1.
DR Pfam: PF00998; Viral RdRp; 1.
DR PrcDom: PD185062; HCV_NS1; 1.
DR SMART: SW00497; DEXDC; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 327466 MW; 4613744EC4DA013 CRC64;

Query Match 95.9%; Score 15356; DB 12; Length 3010;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 2861; Conservatives 69; Mismatches 80; Indels 0; Gaps 0;

QY 1 MSTNPKPQRTKRTNRRPQDVKPPGGQIVGGVYLLPRRGRPRGLGVATRAKASERSQPRG 60
DB 1 MSTNPKPQRTKRTNRRPQDVKPPGGQIVGGVYLLPRRGRPRGLGVATRAKTSERSQPRG 60
QY 61 RRQPIKARPEGRMAQPGYWPFLYNGEGLGWGLLSPRGSPPSWGPTDPRRRSNL 120
DB 61 RRQPIKARPEGRMAQPGYWPFLYNGEGLGWGLLSPRGSPPSWGPTDPRRRSNL 120
QY 121 KVIDTLTCGFADLMGYIPLVGAP;GGARALAHGVRLDGVNATGNLPCSCSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAP;GGARALAHGVRLDGVNATGNLPCSCSIFLLA 180
QY 181 LLSCLTTPASAYEVRNYSIGYHVTNDCSNSSIVYEADVINMHTPCVPCVQEGNSSRCW 240
DB 181 LLSCLTTPASAYEVRNYSIGYHVTNDCSNSSIVYEADVINMHTPCVPCVQEGNSSRCW 240
QY 241 ALTPTLAARNASVETIRRHVDLLVGTAAFCSAMYGDLGCSIPLYSQLETPRRHET 300
DB 241 ALTPTLAARNISVETIRRHVDLLVGTAAFCSAMYGDLGCSFVLSQLETPRRHET 300
QY 301 VQDCNCSLYPGHSGHRMAMWMMNSPTTALVVSQLLRIQAVVDMVAGAHGWGLAGLA 360
DB 301 VQDCNCSLYPGHSGHRMAMWMMNSPTTALVVSQLLRIQAVVDMVAGAHGWGLAGLA 360
QY 361 YYSMVGNWAKVLIIVALLFAGVDGETHTTGRVAGHTTSGTSLFSGASQKIQLVNTNGSW 420
DB 361 YYSMVGNWAKVLIIVMLLFAGVDGTHVTGGTQAARSAILTSFLTVGPSQKIQLVNTNGSW 420
QY 421 HINRTALNCNDSIQTFPAALFYAHKFNSSCPREMASCRPIDMFAQGWGPIITYKPNSS 480
DB 421 HINRTALNCNDSIQTFPAALFYAHKFNSSCPREMASCRPIDMFAQGWGPIITYAEPNNP 480
QY 481 DQRPYCHYAPRCPGWASQVCGPVYCTFSPVVGVTDSRGPVTSWGENETDVMLLN 540
DB 481 DQRPYCHYAPRCPGLEBASQVCGPVYCTFSPVVGVTDRGVPVYKNGENETDVMLLN 540
QY 541 NTRPPQGNWFGTWMNSTGTTCGGPPNCTCGGPNCTGGVGNRTLICPTDFCRKHEPATYTKCGSG 600
DB 541 NTRPPQGNWFGTWMNSTGTTCGGPPNCTCGGPNCTGGVGNRTLICPTDFCRKHEPATYTKCGSG 600
QY 601 PMLTPRCVLDYRLLWHVPCNTLNSIFKVMYVGVGVEHRLNAAQNTWGERONLDDRDS 660
DB 601 PMLTPRCVLDYRLLWHVPCNTVNFTEFKVMYVGVGVEHRLSAACNWTGERCDLDDRDS 660
QY 661 ELSPLLSTTEWQILPCAPFTTLPALSTGLIHLHQNIVDVQYLYGVGSFVSPAIKWEYIL 720
DB 661 ELSPLLSTTEWQILPCSFPTTLPALSTGLIHLHRNIVDVQYLYGVGSFVSPAIKWEYIL 720
QY 721 LIFLLIADARVACACLMMMLLIAQAFAALENLVVLNAAASVAGAHGILSFLVFFCAAAYIKG 780

```

```

721 LIFLLIADARVACACLMMMLLIAQAFAALENLVVLNAAASVAGAHGILSFLVFFCAAAYIKG 780
781 RIAPGAAVAFYGVWPLLLILLIALLPRAVALDREMAASCGGAVLGLVFTLSPYKVFILT 840
781 RLVPGAAYALYGVWPLLLILLIALLPRAVALDREMAASCGGAVFVGLAFITLSPYKVFILA 840
841 RLTIWMLQYFETTRAEAHQVWVPLNVGRGGRDAIILLTCAVHPDLFDITKLLAILGPLM 900
841 RLTIWMLQYFETTRAEAHQVWVPLNVGRGGRDAIILLMVCVHPDLFDITKLLAILGPLM 900
901 VLOAGITRVPYFVRAGLIRACMLVRKVAGHYVQWVFKLKGALICTYTYNHLTPLRDWA 960
901 VLOAGITRVPYFVRAGLIRACMLVRKVAGHYVQWVFKLKAALTCTYTYNHLAPLQHWA 960
961 HAGRLDAVABEVPVFSAMETKVTWGAADTAACGDIILGLPVSARGKEIFLGPADSLG 1020
961 HSGRLDAVABEVPVFSAMETKVTWGAADTAACGDIILGLPVSARGRELLSGPADSLG 1020
1021 QGWELLAPITAXSQQTRGVIGCIITSLGRDNKQVEGEVQVWSTATQSFATCINGVCWT 1080
1021 QGWELLAPITAXSQQTRGVIGCIITSLGRDNKQVEGEVQVWSTATQSFATCINGVCWT 1080
1081 VHGAGSKTLAGPKGPIITQMYTNVDLDLVGWQAPPGARSMPTFCSSGSDLYLVTRHADVI 1140
1081 VFHGAGSKTLAGPKGPIITQMYTNVDQDLVGWQAPPGARSMPTCTCGSSDLYLVTRHADVI 1140
1141 PVRRGDSGSLSPRPVSVLXGSSGPIICPSGHVGVFRAAVCTRGVAKAVDFIPVES 1200
1141 PVRRGDSGSLSPRPVSVLXGSSGPIICPSGHAVGIFRAAVCTRGVAKAVDFIPVES 1200
1201 VETMRSPTDNTSTPPAVPQTPQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVNLPSVAA 1260
1201 VETMRSPTDNTSTPPAVPQTPQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVNLPSVAA 1260
1261 TLGFGYMSXAHGIDENIRGVTFTTIGSITTYTKKFLADCGCGGAYDIIICDECHS 1320
1261 TLGFGYMSXAYGTDNIRGVTFTTIGAPITTYTKKFLADCGCGGAYDIIICDECHS 1320
1321 TDSITLIGITGVLDQAEATAGARLVLAATATPPGSVTVPHPNIBEIGLSNNGNEIPFYKAI 1380
1321 TDSITLIGITGVLDQAEATAGARLVLAATATPPGSVTVPHPNIBEVALSNTGEIPFYKAI 1380
1381 PIBAKGRHLIFCHSKKCKDELAALKLGLNNAVAYRGLDVSVIPIIDGVVVVATDAL 1440
1381 PIEVIGKGRHLIFCHSKKCKDELAALKLGLNNAVAYRGLDVSVIPIIDGVVVVATDAL 1440
1441 MTGPTGDFSDVIDCNTCTVQTVDFSLDPTFTTITVTPQDAVSRQSRGRTGRSGIYR 1500
1441 MTGPTGDFSDVIDCNTCTVQTVDFSLDPTFTTITVTPQDAVSRQSRGRTGRSGIYR 1500
1501 FVTPGERPSPGMFSSVLCVCEVDAGCAWYELTPAETSVELRAYLNTPLGVCODHLEFWES 1560
1501 FVTPGERPSPGMFSSVLCVCEYDAGCAWYELTPAETSVELRAYLNTPLGVCODHLEFWES 1560
1561 VFTGLTHIDAHFTSQTKQAGDNFPYLVAQATVCARAQAPPPSWDQWKKCLIRLKPTLHG 1620
1561 VFTGLTHIDAHFTSQTKQAGDNFPYLVAQATVCARAQAPPPSWDQWKKCLIRLKPTLHG 1620
1621 PTPLLYRLGAVQNEVILTHPTTKYIMACMSADLEVNTSTWLVGVJLAALAAVCLITGVS 1680
1621 PTPLLYRLGAVQNEVILTHPTTKYIMACMSADLEVNTSTWLVGVJLAALAAVCLITGVS 1680
1681 VVGRITILSGKPAVVPDREVLQSEFDEMERCASQPIEQGMQLAEQFKQKALGLLTAT 1740
1681 VVGRITILSGKPAVVPDREVLQSEFDEMERCASQPIEQGMQLAEQFKQKALGLLTAT 1740
1741 KOAFAAAPVSVESKWRALETFWAKHWNFIISQIYLAGLSTLPGNPALASMAFTASTSP 1800
1741 KOAFAAAPVSVESKWRALETFWAKHWNFIISQIYLAGLSTLPGNPALASMAFTASTSP 1800
1801 LTTQNTLLFNLTLCGMVAQAQAPPSSAAAFVAGIAGAAVSGIGLGVLDILAGYAGVA 1860

```

Db 1801 LTTQTTLMFNILGWAAQAAPPASAFYAGAGIAGAAGVSGIGLKVLDILAGYAGVA 1860
QY 1861 GALVAFVMSGEVFESTBDLVNLLPAILSPALGVGVVCAAILRRHVGPBGGAQVMNRLI 1920
Db 1861 GALVAFVMSGEVFESTBDLVNLLPAILSPALGVGVVCAAILRRHVGPBGGAQVMNRLI 1920
QY 1921 AFASRGNEHVSPTHYVPESDAARVTQILSSLTITQLLKRHHQWINEDECSTPCSGMLRDV 1980
Db 1921 AFASRGNEHVSPTHYVPESDAARVTQILSSLTITQLLKRHHQWINEDECSTPCSGMLRDV 1980
QY 1981 WDMICTVLTDFKWLQSKLPRLPBGVFLSCORGKGVWGDGTMTQTCGQIAGHVK 2040
Db 1981 WDMICTVLSDFKWLQSKIMPRIPGVFLSCORGKGVWGDGTMTQTCGQIAGHVK 2040
QY 2041 NGSMRIIVGPTCSNTHGTTPFINAYTGPCTPSPAPNYSRALRVAAEEYVETRVGDFH 2100
Db 2041 NGSMRIIVGPTCSNTHGTTPFINAYTGPCTPSPAPNYSRALRVAAEEYVETRVGDFH 2100
QY 2101 YVTGMTTNDVKPCQVPAPPEFFTEVDGVRHLRYAPACKPLLEEDVTPQVGLNQYLVGSQL 2160
Db 2101 YVTGMTTNDVKPCQVPAPPEFFTEVDGVRHLRYAPACKPLLEEDVTPQVGLNQYLVGSQL 2160
QY 2161 PCPEPDDVTLTSMITDPSHITAEAKRLARGSPPSLASSASQISAPSKATCTTHD 2220
Db 2161 PCPEPDDVTLTSMITDPSHITAEAKRLARGSPPSLASSASQISAPSKATCTTHD 2220
QY 2221 SPDALEALLMRQMGNIIVSESNKVILDSFEPLAEAGDEREISVAAEILNRKSRK 2280
Db 2221 SPDALEALLMRQMGNIIVSESNKVILDSFEPLAEAGDEREISVAAEILNRKSRK 2280
QY 2281 FPALPTWAPDYNPPLLESKDPDYVPPVHGCPTLPPTXAPPIPPRRKRTVVLTESNV 2340
Db 2281 FPALPTWAPDYNPPLLESKDPDYVPPVHGCPTLPPTXAPPIPPRRKRTVVLTESNV 2340
QY 2341 SSALAEALATKTFSGSSAVDSGTATLPLASDDGDGKSDVSESYSMPPLEGEPPDL 2400
Db 2341 SSALAEALATKTFSGSSAVDSGTATLPLASDDGDGKSDVSESYSMPPLEGEPPDL 2400
QY 2401 SDGWSVTSEASEDVEDVCCSMSTWTGALITPCAABESKLPINPLNSLRHNNVYATT 2460
Db 2401 SDGWSVTSEASEDVEDVCCSMSTWTGALITPCAABESKLPINPLNSLRHNNVYATT 2460
QY 2461 SRASLRQKVTDFRLQVLDHVDVLEKMKAKASTVAKALLIEBACKLTPPHSARSKF 2520
Db 2461 SRASLRQKVTDFRLQVLDHVDVLEKMKAKASTVAKALLIEBACKLTPPHSARSKF 2520
QY 2521 GYGAKDVRLSSRAVNHRSVWEDLLEDTETPTDITIMAKSEVFCVQPEKGRKPARLIV 2580
Db 2521 GYGAKDVRLSSRAVNHRSVWEDLLEDTETPTDITIMAKSEVFCVQPEKGRKPARLIV 2580
QY 2581 FPDGLVRVCEKMALYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTWKSCKCPMGFSYDT 2640
Db 2581 FPDGLVRVCEKMALYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTWKSCKCPMGFSYDT 2640
QY 2641 RCFDSTVTESDIRVESIYOCCLAPAROAIRSLTERLVIIGPLTNSKQNGCYGRORA 2700
Db 2641 RCFDSTVTESDIRVESIYOCCLAPAROAIRSLTERLVIIGPLTNSKQNGCYGRORA 2700
QY 2701 SGVLTTCGNTLTCYLKATAACRAKLQDCTMLVNGDLDLVIICESAGTOEDAAALRAFT 2760
Db 2701 SGVLTTCGNTLTCYLKATAACRAKLQDCTMLVNGDLDLVIICESAGTOEDAAALRAFT 2760
QY 2761 AMTRYAPPDPPQPEYDLELITSCSNVSVAHDSKRVYVLTPTTFLAPAAWETAR 2820
Db 2761 AMTRYAPPDPPQPEYDLELITSCSNVSVAHDSKRVYVLTPTTFLAPAAWETAR 2820
QY 2821 HTPINSLWLGNIIVAPTLWARMILMTHFRSILLAQQLKALCOIYGACYSTEPDLDPQ 2880
Db 2821 HTPINSLWLGNIIVAPTLWARMILMTHFRSILLAQQLKALCOIYGACYSTEPDLDPQ 2880
QY 2881 IIBELHGLSAFTLHSHYSPGEBINRVASCLRLKGVPLRLTWHRARSVRKLLSOGGAATC 2940
Db 2881 IIBELHGLSAFTLHSHYSPGEBINRVASCLRLKGVPLRLTWHRARSVRKLLSOGGAATC 2940

QY 2941 GRYLFNWAVERTKLTPIPAASOLDLSGWFVAGYSGGDIYHSLSRARPRWFPCLLLLSV 3000
Db 2941 GRYLFNWAVERTKLTPIPAASOLDLSGWFVAGYSGGDIYHSLSRARPRWFPCLLLLSV 3000
QY 3001 GVGIYLLPNR 3010
Db 3001 GVGIYLLPNR 3010
RESULT 12
Q9J3H9 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H9; (T=EMBLrel. 15, Created)
DT 01-OCT-2000 (T=EMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (T=EMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (T=EMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD13;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
progression."
RL Submitted (NCV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF207754; AAF65944.1;
DR PIR: A61196; A61196.
DR PIR: PQ0246; PQ0246.
DR PIR: PQ0329; PQ0329.
DR HSSP: P26663; IJXP.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0019028; C:viral capsid; IEA.
DR GO: GO:0019031; C:viral envelope; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
DR GO: GO:0001938; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR GO: GO:0019079; P:viral genome replication; IEA.
DR GO: GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR000345; Cys_Ser_trypsin.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR000745; HCV NS4a.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RDRP.
DR InterPro: IPR001650; Helicase C.
DR InterPro: IPR004109; Peptidase C29.
DR InterPro: IPR007095; RNA pol DS PS.
DR InterPro: IPR007094; RNA pol PSvir.
DR Pfam: PF01543; HCV capsid; 1.
DR Pfam: PF01542; HCV core; 1.
DR Pfam: PF01539; HCV env; 1.

Qy	1921	AFASRGNHVS	THYVP	SDAAR	VTQ	LISS	LT	ITQL	KRI	HOW	IN	EDCS	TP	CSGS	WLRD	V	1980																																													
Db	1921	AFASRGNHVS	THYVP	SDAAR	VTQ	LISS	LT	ITQL	KRI	HOW	IN	EDCS	TP	CSGS	WLRD	V	1980																																													
Qy	1981	WDMCT	VLTD	FKTML	QSK	L	PRL	PG	VP	PLSC	Q	RG	YK	GV	WR	GDG	IMQTT	PC	GAQ	IAGHV	K	2040																																								
Db	1981	WDMCT	VLAD	FKAML	QSK	L	PRL	PG	VP	PLSC	Q	RG	YK	GV	WR	GDG	IMQTT	PC	GAQ	I	THVK	K	2040																																							
Qy	2041	NGS	RI	V	UG	PR	TCS	N	WH	G	T	P	I	N	A	Y	T	G	P	C	T	P	S	2100																																						
Db	2041	NGS	RI	V	UG	PR	TCS	N	WH	G	T	P	I	N	A	Y	T	G	P	C	T	P	S	2100																																						
Qy	2101	YVT	G	M	T	D	N	V	K	P	C	Q	V	P	A	P	E	F	T	E	V	D	G	V	2160																																					
Db	2101	YVT	G	M	T	D	N	V	K	P	C	Q	V	P	A	P	E	F	T	E	V	D	G	V	2160																																					
Qy	2161	PCE	PE	P	D	V	I	L	T	S	M	L	T	D	P	S	H	I	T	A	E	T	A	K	R	2220																																				
Db	2161	PCE	PE	P	D	V	A	V	T	S	M	L	T	D	P	S	H	I	T	A	E	T	A	K	R	2220																																				
Qy	2221	SPD	AD	L	E	AN	L	W	R	Q	E	M	G	N	T	R	V	E	S	E	N	K	V	I	L	S	2280																																			
Db	2221	SPD	AD	L	E	AN	L	W	R	Q	E	M	G	N	T	R	V	E	S	E	N	K	V	I	L	S	2280																																			
Qy	2281	FPS	AL	P	I	W	A	R	D	Y	N	P	P	L	E	S	W	K	D	P	D	Y	V	P	V	V	2340																																			
Db	2281	FPS	AL	P	I	W	A	R	D	Y	N	P	P	L	E	S	W	K	D	P	D	Y	V	P	V	V	2340																																			
Qy	2341	SSA	LA	E	L	A	T	K	T	F	G	S	S	A	V	D	S	G	T	A	T	A	L	P	O	A	S	2400																																		
Db	2341	SSA	LA	E	L	A	T	K	T	F	G	S	S	A	V	D	S	G	T	A	T	A	L	P	O	A	S	2400																																		
Qy	2401	SDG	S	N	T	S	V	E	B	A	S	D	V	V	C	C	S	N	Y	T	W	T	G	A	L	I	P	C	A	A	E	S	K	L	P	I	N	F	L	S	N	L	L	R	H	N	N	V	A	T	2460											
Db	2401	SDG	S	N	T	S	V	E	B	A	S	D	V	V	C	C	S	N	Y	T	W	T	G	A	L	I	P	C	A	A	E	S	K	L	P	I	N	F	L	S	N	L	L	R	H	N	N	V	A	T	2460											
Qy	2461	SRS	AS	L	E	O	K	V	T	F	D	R	I	O	V	L	D	H	Y	D	V	L	K	E	M	K	A	K	A	S	T	W	A	K	L	L	S	T	E	E	A	C	K	T	P	P	H	S	A	K	S	F	2520									
Db	2461	SRS	AS	L	E	O	K	V	T	F	D	R	I	O	V	L	D	H	Y	D	V	L	K	E	M	K	A	K	A	S	T	W	A	K	L	L	S	T	E	E	A	C	K	T	P	P	H	S	A	K	S	F	2520									
Qy	2521	G	Y	K	A	K	O	V	E	N	L	S	S	R	A	V	N	H	I	R	S	W	E	D	L	I	B	E	T	P	I	D	T	I	T	I	M	A	K	S	E	P	C	V	O	P	E	K	G	R	K	P	A	R	L	I	V	2580				
Db	2521	G	Y	K	A	K	O	V	E	N	L	S	S	R	A	V	N	H	I	R	S	W	E	D	L	I	B	E	T	P	I	D	T	I	T	I	M	A	K	S	E	P	C	V	O	P	E	K	G	R	K	P	A	R	L	I	V	2580				
Qy	2581	F	P	D	L	G	V	R	C	K	M	A	L	Y	D	V	V	S	T	L	P	O	A	V	M	S	S	Y	G	F	O	Y	S	P	K	O	R	V	E	L	V	N	T	W	K	S	K	C	P	M	G	F	S	Y	D	T	2640					
Db	2581	F	P	D	L	G	V	R	C	K	M	A	L	Y	D	V	V	S	T	L	P	O	A	V	M	S	S	Y	G	F	O	Y	S	P	K	O	R	V	E	L	V	N	T	W	K	S	K	C	P	M	G	F	S	Y	D	T	2640					
Qy	2641	R	C	F	D	S	T	V	T	S	D	I	R	V	E	S	I	O	C	C	D	L	A	P	A	R	O	A	I	R	S	L	T	E	R	L	I	G	P	L	T	N	S	K	O	N	C	O	N	Y	R	C	R	A	2700							
Db	2641	R	C	F	D	S	T	V	T	S	D	I	R	V	E	S	I	O	C	C	D	L	A	P	A	R	O	A	I	R	S	L	T	E	R	L	I	G	P	L	T	N	S	K	O	N	C	O	N	Y	R	C	R	A	2700							
Qy	2701	S	G	V	L	T	S	C	G	N	T	L	A	C	A	C	R	A	K	L	O	D	C	T	M	L	V	N	G	D	L	V	I	C	E	S	A	G	T	O	R	D	E	A	A	L	A	F	T	E	2760											
Db	2701	S	G	V	L	T	S	C	G	N	T	L	A	C	A	C	R	A	K	L	O	D	C	T	M	L	V	N	G	D	L	V	I	C	E	S	A	G	T	O	R	D	E	A	A	L	A	F	T	E	2760											
Qy	2761	A	M	T	R	S	A	P	P	G	D	P	O	P	E	V	D	L	E	L	I	T	S	C	S	N	S	V	A	H	D	A	S	G	R	V	Y	L	T	R	D	T	T	P	L	A	R	A	A	E	T	A	R	2820								
Db	2761	A	M	T	R	S	A	P	P	G	D	P	O	P	E	V	D	L	E	L	I	T	S	C	S	N	S	V	A	H	D	A	S	G	R	V	Y	L	T	R	D	T	T	P	L	A	R	A	A	E	T	A	R	2820								
Qy	2821	H	T	P	N	S	W	L	G	N	I	M	A	P	T	L	W	A	R	M	I	N	T	H	F	F	S	I	L	A	O	E	L	A	C	O	I	Y	G	A	C	S	I	B	P	D	L	P	O	2880												
Db	2821	H	T	P	N	S	W	L	G	N	I	M	A	P	T	L	W	A	R	M	I	N	T	H	F	F	S	I	L	A	O	E	L	A	C	O	I	Y	G	A	C	S	I	B	P	D	L	P	O	2880												
Qy	2881	I	I	E	R	L	H	G	L	S	A	F	T	L	H	S	Y	S	P	G	E	I	N	R	V	A	S	C	L	R	K	L	G	I	G	V	P	P	L	T	W	E	R	H	R	A	R	S	V	R	A	K	L	S	O	G	G	R	A	T	C	2940
Db	2881	I	I	E	R	L	H	G	L	S	A	F	T	L	H	S	Y	S	P	G	E	I	N	R	V	A	S	C	L	R	K	L	G	I	G	V	P	P	L	T	W	E	R	H	R	A	R	S	V	R	A	K	L	S	O	G	G	R	A	T	C	2940
Qy	2941	G	R	V	L	F	N	W	A	V	T	K	L	T	S	-	P	A	A	S	O	L	D	S	G	W	F	V	A	G	S	G	D	I	V	H	S	L	S	R	A	P	P	R	F	P	L	C	L	L	L	S	3000									
Db	2941	G	K	V	L	F	N	W	A	V	T	K	L	T	S	-	P	A	A	S	O	L	D	S	G	W	F	V	A	G	S	G	D	I	V	H	S	L	S	R	A	P	P	R	F	P	L	C	L	L	L	S	3000									

```

Qy 3001 GVGIVLLPNR 3010
Db 3001 GVGIVLLPNR 3010
|||||
|||||
|||||

RESULT 13
P89966 PRELIMINARY; PRT; 3010 AA.
AC P89966;
DT 01-MAY-1997 (rEMBLrel. 03, Created)
DT 01-MAY-1997 (rEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (rEMBLrel. 25, Last annotation update)
DE RNA for polyprotein (Genome polyprotein).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=type 1b;
RA Tanaka T.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
[2];
RP SEQUENCE FROM N.A.
RC STRAIN=type 1b;
RA TANAKA T.;
RT "MORE.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL; D89872; BAAL4035.1; -.
DR PIR; A61196; A61196.
DR PIR; PQ0245; PQ0246.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0329; PS0329.
DR HSP; P26663; LUXP.
DR GQ; GQ:0016021; C:integral to membrane, IEA.
DR GQ; GQ:0013028; C:viral capsid, IEA.
DR GQ; GQ:0013031; C:viral envelope, IEA.
DR GQ; GQ:0005524; E:ATP binding, IEA.
DR GQ; GQ:0008026; F:ATP dependent helicase activity; IEA.
DR GQ; GQ:0003723; F:RNA binding, IEA.
DR GQ; GQ:0003969; F:RNA-directed RNA polymerase activity; IEA.
DR GQ; GQ:0008236; F:serine-type peptidase activity; IEA.
DR GQ; GQ:0005198; F:structural molecule activity; IEA.
DR GQ; GQ:0016749; E:transferase activity; IEA.
DR GQ; GQ:0006508; F:proteolysis and peptidolysis; IEA.
DR GQ; GQ:0006350; F:transcription, IEA.
DR GQ; GQ:0013079; F:viral genome replication, IEA.
DR GQ; GQ:0019087; F:viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002521; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_C29.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PsIVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.

```


QY 1981 WDWICTVLTDEKTLQSKLLPRLPGVPELSGORGKGVWRGDMQTTCPGCAQIAGHVK 2040
 Db 1981 WDWICTVLTDEKTLQSKLLPRLPGVPELSGORGKGVWRGDMQTTCPGCAQIAGHVK 2040
 QY 2041 NGSMEIVGERTCSNTWHTGTFINAYTTGCTPSPAPNYSRALRVAABEYVEVTRVGDFH 2100
 Db 2041 NGSMEIVGERTCSNTWHTGTFINAYTTGCTPSPAPNYSRALRVAABEYVEVTRVGDFH 2100
 QY 2101 YVTGTTNDVAKPCQVPAPEFTEVDGVRLLRYAPACPLLEDVTFQVGLNQVLVGSOL 2160
 Db 2101 YVTGTTNDVAKPCQVPAPEFTEVDGVRLLRYAPACPLLEDVTFQVGLNQVLVGSOL 2160
 QY 2161 PCEPPEPDVAVLTSMULTDSHITAETAKRLARGSPSLASSASOLSAPLSKATCTTHD 2220
 Db 2161 PCEPPEPDVAVLTSMULTDSHITAETAKRLARGSPSLASSASOLSAPLSKATCTTHD 2220
 QY 2221 SPDAADLIENLWRQEMGNITRVSEKNVWILDSFPLRAEEDEREVSVAEILLRKSXK 2280
 Db 2221 SPDAADLIENLWRQEMGNITRVSEKNVWILDSFPLRAEEDEREVSVAEILLRKSXK 2280
 QY 2281 FESALPIWARPDYNPELLESKWDPPYVPPVHGCPLPTKAPPIDPPRKRVTVLTBSNV 2340
 Db 2281 FESALPIWARPDYNPELLESKWDPPYVPPVHGCPLPTKAPPIDPPRKRVTVLTBSNV 2340
 QY 2341 SSALAEATKTFGSGSAVDSGTATLPDLASDDGXGDSVESYSSMPPLEGEGDPDL 2400
 Db 2341 SSALAEATKTFGSGSAVDSGTATLPDLASDDGXGDSVESYSSMPPLEGEGDPDL 2400
 QY 2401 SDGWSVTSEBSEDVVCSSSYTWGTALITPCAARESKLPINPLNSLLRHNQVYATT 2460
 Db 2401 SDGWSVTSEBSEDVVCSSSYTWGTALITPCAARESKLPINPLNSLLRHNQVYATT 2460
 QY 2461 SRSASLRQKVTDFDLQVLDHRYDLKEMKAKASTVKAKLLSBEACKLTPPHSAKSKF 2520
 Db 2461 SRSASLRQKVTDFDLQVLDHRYDLKEMKAKASTVKAKLLSBEACKLTPPHSAKSKF 2520
 QY 2521 GYGAKDVRLNLSRAVNHRSVWEDLLEDTETPIDTINAKSEVFCVQEKGRKPARLIV 2580
 Db 2521 GYGAKDVRLNLSRAVNHRSVWEDLLEDTETPIDTINAKSEVFCVQEKGRKPARLIV 2580
 QY 2581 FPDGLGVRCERKALYDVVSTLTPQAVMGSSYGFQYSPKQVRFLVNTWKSCKCPMGFSYDT 2640
 Db 2581 FPDGLGVRCERKALYDVVSTLTPQAVMGSSYGFQYSPKQVRFLVNTWKSCKCPMGFSYDT 2640
 QY 2641 RCFDSTVTESDIRVEESTYQCCDLAPAEARQALRSILTERLYIGGLTNSKGQCGYRCRA 2700
 Db 2641 RCFDSTVTESDIRVEESTYQCCDLAPAEARQALRSILTERLYIGGLTNSKGQCGYRCRA 2700
 QY 2701 SGVLTTSCTNTLTCLYKATAACRAAKLQDCTMLVNGDDLVIKESAGTQEDAAALRAFT 2760
 Db 2701 SGVLTTSCTNTLTCLYKATAACRAAKLQDCTMLVNGDDLVIKESAGTQEDAAALRAFT 2760
 QY 2761 ANTRYSAPEGPPQPEYDLLEITSCSSNVVAHDASGRVYILTRDPTTPARAWEVAR 2820
 Db 2761 ANTRYSAPEGPPQPEYDLLEITSCSSNVVAHDASGRVYILTRDPTTPARAWEVAR 2820
 QY 2821 HTPINSWLGNIIMVAPTLMWMLTHFFSVILLAQEQLKALDQOYGCACYSIRPLDLPQ 2880
 Db 2821 HTPINSWLGNIIMVAPTLMWMLTHFFSVILLAQEQLKALDQOYGCACYSIRPLDLPQ 2880
 QY 2881 IIERLHGLSAPLSHSYSGEINRVASCLKLGVPPLRTWRHRSVRKALLSQGGRATC 2940
 Db 2881 IIERLHGLSAPLSHSYSGEINRVASCLKLGVPPLRTWRHRSVRKALLSQGGRATC 2940
 QY 2941 GRYLENNAVRKLTLPAPASOLDLSGFVAGYSGGDIYHLSRARPWPPLCLLLLSV 3000
 Db 2941 GRYLENNAVRKLTLPAPASOLDLSGFVAGYSGGDIYHLSRARPWPPLCLLLLSV 3000
 QY 3001 GVGIVLLPNR 3010
 Db 3001 GVGIVLLPNR 3010

RESULT 14
 Q9DTE6 PRELIMINARY; PRT; 3010 AA.
 AC Q9DTE6;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCVTL142;
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
 RA Takahashi T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
 RA Mishiro S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 RT with hepatocellular carcinoma: the 'progression score' revisited.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MENA (BY SIMILARITY).
 CC EXBL; AB049031; BAB18804.1; -.
 DR PIR; A61196; A61196.
 DR HSP; P26663; LUXP.
 DR CO; GO:0016021; C:integral to membrane; IEA.
 DR CO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0004899; F:electron transporter activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003958; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0005198; F:serine-type peptidase activity; IEA.
 DR GO; GO:0016740; F:structural molecule activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro; IPR000345; Cys_Ser_trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.

DR	Pfam; PF00998; Viral_rdrp; 1.	
DR	ProDom; PD186062; HCV NS1; 1.	
DR	SMART; SM00487; DEXDc; 1.	
DR	SMART; SM00490; HELICg; 1.	
DR	PROSITE; PS00190; CYTOCHROME_C; 1.	
KW	Arp-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;	
KW	Hydrolase; Nonstructural protein; Polypeptide;	
KW	RNA-directed RNA polymerase; Transferase; Transmembrane.	
SQ	SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;	
Query Match		
Best Local Similarity 95.8%; Score 15330; DB 12; Length 3010;		
Matches 2850; Conservative 75; Mismatches 85; Indels 0; Gaps 0;		
QY	1 MSTNPKPQKTKNTNRRPQDKVFGGQIVGGVILLPRGPRGLGVRAURKASERSQPRG 60	
DB	1 MSTNPKPQKTKNTNRRPQDKVFGGQIVGGVILLPRGPRGLGVRAURKASERSQPRG 60	
QY	61 RQOPTKARPRGRWAQPGYWPPLYVGNELGWAGMLLSRGRSPSGWPTDPRRSNIG 120	
DB	61 RQOPTKARPRGRWAQPGYWPPLYVGNELGWAGMLLSRGRSPSGWPTDPRRSNIG 120	
QY	121 KVIDLTCGFADLMGYIPLVGLAPLGAARALAHGVRLSDGVNYATGNLPGCSFSIFLLA 180	
DB	121 KVIDLTCGFADLMGYIPLVGLAPLGAARALAHGVRLSDGVNYATGNLPGCSFSIFLLA 180	
QY	181 LLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYEAADVIMHTPGVPCVQBGNSRCWV 240	
DB	181 LLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYEAADVIMHTPGVPCVQBGNSRCWV 240	
QY	241 ALTPILAAASVPTTIRHVDLLVGTAAAPCSAMYVGLCGSIFLVSQLFTFSPRRHET 300	
DB	241 ALTPILAAASVPTTIRHVDLLVGTAAAPCSAMYVGLCGSIFLVSQLFTFSPRRHET 300	
QY	301 VDCNCSIYPGVSHGRXAWDMNMSPTALVWSOLLRIPOAVDMVAGAHGVLGALA 360	
DB	301 VDCNCSLYPGVSHGRXAWDMNMSPTALVWSOLLRIPOAVDMVAGAHGVLGALA 360	
QY	361 YYSVMGNWAKVILVALLFAGVDGETHTGVRAGHTTSGTSLFSSGASOKIQLVNTINGSW 420	
DB	361 YYSVMGNWAKVILVALLFAGVDGETHTGVRAGHTTSGTSLFSSGASOKIQLVNTINGSW 420	
QY	421 HINRTALNCNDSLOTCFFAALFYAHKZNSGCGPERWASCRPIDWFAQAGPITYIKPNS 480	
DB	421 HINRTALNCNDSLOTCFFAALFYAHKZNSGCGPERWASCRPIDWFAQAGPITYIKPNS 480	
QY	481 DQRPYCHWYAPPCGVVPVQASQVGPYCTPSPVVGTTDRSGVPTYSNGENETDVLIN 540	
DB	481 DQRPYCHWYAPPCGVVPVQASQVGPYCTPSPVVGTTDRSGVPTYSNGENETDVLIN 540	
QY	541 NTRPQGNWFGCTWNNSTGFTKTCGGPPCNIIGVGNRTLICPTDCFRKHPEATYTKCGSG 600	
DB	541 NTRPQGNWFGCTWNNSTGFTKTCGGPPCNIIGVGNRTLICPTDCFRKHPEATYTKCGSG 600	
QY	601 PMLTPRELVDPYRLWHYPTLNFSIFKVMYVGVGVEHRLNACNWRGRCNLEDRBS 660	
DB	601 PMLTPRELVDPYRLWHYPTLNFSIFKVMYVGVGVEHRLNACNWRGRCNLEDRBS 660	
QY	661 ELSPLLSTTEWQILPCAPTTLPALSTGLIHLHQNIVDVQVLYGVGSAVSPFAIKWEYL 720	
DB	661 ELSPLLSTTEWQILPCAPTTLPALSTGLIHLHQNIVDVQVLYGVGSAVSPFAIKWEYL 720	
QY	721 LLFLLADARVCACLMWLLIIQAEAALENLVVINAASVAGAGHILSFIVFFCAAWYIKG 780	
DB	721 LLFLLADARVCACLMWLLIIQAEAALENLVVINAASVAGAGHILSFIVFFCAAWYIKG 780	
QY	781 RLAPGAAVAFYGVWPLLILLALPPAYALDREMAASCGGAVLVGLVFTLSPYKVFELT 840	
DB	781 KLVPGRAYAFYGVWPLLILLALPPAYALDREMAASCGGAVLVGLVFTLSPYKVFELT 840	
QY	841 RLIIWLQYFITRAEAHQVWVPLNVGRGDAIILLTCAVHPELIFDITKLLAILLGLPLM 900	
DB	841 KLIIWLQYFITRAEAHQVWVPLNVGRGDAIILLTCAVHPELIFDITKLLAILLGLPLM 900	

QY	901 VLOAGITRVPYFVRACGLIRACMLVRKVAGGHYVQVFMKLCALGTGTVVYVNHLEPLRDA 960	
DB	901 VLOAGITRVPYFVRACGLIRACMLVRKVAGGHYVQVFMKLCALGTGTVVYVNHLEPLRDA 960	
QY	961 HAGLRLDAVAVPEVPVFSAMETKVTWADTAACGDIILGLPVSARRKEIFLGPADSLEG 1020	
DB	961 HTGLRLDAVAVPEVPVFSAMETKVTWADTAACGDIISGLPVSARRKEIFLGPADSLEG 1020	
QY	1021 GQWRLLAPITAYSQOTRGVLGCIITSLTRDKNOVEGEVQVUSTATOSFLATCTNGVCWT 1080	
DB	1021 GQWRLLAPITAYSQOTRGVLGCIITSLTRDKNOVEGEVQVUSTATOSFLATCTNGVCWT 1080	
QY	1081 VYHAGSXTLAGPKPIQMYTNVDLVLGWAQAPPGARSMTPCSCGSDLYLVTRHADVI 1140	
DB	1081 VYHAGSXTLAGPKPIQMYTNVDLVLGWAQAPPGARSMTPCSCGSDLYLVTRHADVI 1140	
QY	1141 PVRRRGDSRGSLLSPRPVSYLKGSGGGLLCPSGHVVGIFFAAVCTRGVAKAVDPVPEA 1200	
DB	1141 PVRRRGDSRGSLLSPRPVSYLKGSGGGLLCPSGHVVGIFFAAVCTRGVAKAVDPVPEA 1200	
QY	1201 MTTWRSVFTDNTSPPAVPCOTFOVAHLFAPTGSKSTKPAAYAAOQYKVLVLPNSVAA 1260	
DB	1201 MTTWRSVFTDNTSPPAVPCOTFOVAHLFAPTGSKSTKPAAYAAOQYKVLVLPNSVAA 1260	
QY	1261 TLGPGAYMSKAGHDIPNIRTGVRTITTGSIITYSTYGHFLADGGCGGAYDIIICDECHS 1320	
DB	1261 TLGPGAYMSKAGHDIPNIRTGVRTITTGSIITYSTYGHFLADGGCGGAYDIIICDECHS 1320	
QY	1321 TDSITILGIGVILDOAETAGARLVLAATATPPGVSIVTPHPNIEBGLSNNGEIPFYGKAI 1380	
DB	1321 TDSITILGIGVILDOAETAGARLVLAATATPPGVSIVTPHPNIEBGLSNNGEIPFYGKAI 1380	
QY	1381 PTEALTKGRHLIFCHSKKKCDLAALKLGLGNAYVYRGLDVSVIPIGVDVVVATDAL 1440	
DB	1381 PTEALTKGRHLIFCHSKKKCDLAALKLGLGNAYVYRGLDVSVIPIGVDVVVATDAL 1440	
QY	1441 MTGTGDFSDVIDCNTCTQTVDPSLDPTFTIETTTVPQDAVSRSQRRGTRGRSGIYR 1500	
DB	1441 MTGTGDFSDVIDCNTCTQTVDPSLDPTFTIETTTVPQDAVSRSQRRGTRGRSGIYR 1500	
QY	1501 FVTGERSGMPDSSVLCEDAGCAWYELTPAETSVELRAYLNTPLPGLVPCODHLEFWS 1560	
DB	1501 FVTGERSGMPDSSVLCEDAGCAWYELTPAETSVELRAYLNTPLPGLVPCODHLEFWS 1560	
QY	1561 VFTGLTHIDAFPLSOTKQAGNFYLVAYQATVCARAQAPPPSDQMWKCLIRLKPTLHG 1620	
DB	1561 VFTGLTHIDAFPLSOTKQAGNFYLVAYQATVCARAQAPPPSDQMWKCLIRLKPTLHG 1620	
QY	1621 PTELLYRLGAYQNEVILFHPITKYMCSADLEVVVTSTWLVGSLAALAAYCLITGCV 1680	
DB	1621 PTELLYRLGAYQNEVILFHPITKYMCSADLEVVVTSTWLVGSLAALAAYCLITGCV 1680	
QY	1681 VIVGRIILSGGPAVVPDBEVLYOBDEMECAQSOLPIECGMOLAEOKAKGLQAT 1740	
DB	1681 VIVGRIILSGGPAVVPDBEVLYOBDEMECAQSOLPIECGMOLAEOKAKGLQAT 1740	
QY	1741 KOAEAAAFVWSKWEALFETFWAKMNPISGIQVLAGLSTLPGNPAISLMAFTASTSP 1800	
DB	1741 KOAEAAAFVWSKWEALFETFWAKMNPISGIQVLAGLSTLPGNPAISLMAFTASTSP 1800	
QY	1801 LTTQNTLLFNILGKWWAAQLAPPSAASAFVAGIAGAAVSGIGLGKVLVDILAGYGAVA 1860	
DB	1801 LTTQNTLLFNILGKWWAAQLAPPSAASAFVAGIAGAAVSGIGLGKVLVDILAGYGAVA 1860	
QY	1861 GALVAFKMSGEVSTEDLVNLLPALISPGALVGVVCAAILRRHVHGEGAVOMNRLI 1920	
DB	1861 GALVAFKMSGEVSTEDLVNLLPALISPGALVGVVCAAILRRHVHGEGAVOMNRLI 1920	
QY	1921 AFASRGHVSHTVVPESDAARVTOILSSITITQLLKLHOWINEDCSTPCSGSWLRDV 1980	
DB	1921 AFASRGHVSHTVVPESDAARVTOILSSITITQLLKLHOWINEDCSTPCSGSWLRDV 1980	

```
Qy 1981 WDWICTVLTDKTKLQSKLLPRLPGVPLSQORGKGVWRGDMQTCPCQAGIAGHVK 2040
Db 1981 WDWICTVLTDKTKLQSKLLPRLPGVPLSQORGKGVWRGDMQTCPCQAGIAGHVK 2040
Qy 2041 NGSNRIIVGPRCTCSNTWGTGTPFNINATYTGPCPTSPAPNYSRLMVAABEYVEVTRVGDFF 2100
Db 2041 NGSNRIIVGPRCTCSNTWGTGTPFNINATYTGPCPTSPAPNYSRLMVAABEYVEVTRVGDFF 2100
Qy 2101 YVTGMTTNNVKCPQVPAPEFTEVDGVRLLHRYAPACKPLREDDVTFQVGLNQVLVSQ 2160
Db 2101 YVTGMTTNNVKCPQVPAPEFTEVDGVRLLHRYAPACKPLREDDVTFQVGLNQVLVSQ 2160
Qy 2161 PCEPEPDVTLTSMITDPSHTAETAKERLARGSPPSLASSASQISAPSLKATCTTHD 2220
Db 2161 PCEPEPDVTLTSMITDPSHTAETAKERLARGSPPSLASSASQISAPSLKATCTTHD 2220
Qy 2221 SPDADLIENILNQEMGNITRVESENKVILDSFEPHLHAGBEREISVAATILRSKRK 2280
Db 2221 SPDADLIENILNQEMGNITRVESENKVILDSFEPHLHAGBEREISVAATILRSKRK 2280
Qy 2281 FPSALPTWARPDYAPNPLLESWKDPYVPPVHGCPLEPTKAPPIPPRRKKEVVLTSNV 2340
Db 2281 FPSALPTWARPDYAPNPLLESWKDPYVPPVHGCPLEPTKAPPIPPRRKKEVVLTSNV 2340
Qy 2341 SSALAEATKTFGSSGSAVDSGTATALLPLDASDDGKSDVESYSMPPLLEGPGDPDL 2400
Db 2341 SSALAEATKTFGSSGSAVDSGTATALLPLDASDDGKSDVESYSMPPLLEGPGDPDL 2400
Qy 2401 SDGSKSTVSEAEADVCCNSYWTGALITPCAABESKLPINPLNSLLRHNNVYATT 2460
Db 2401 SDGSKSTVSEAEADVCCNSYWTGALITPCAABESKLPINPLNSLLRHNNVYATT 2460
Qy 2461 SRSASLRQKVTFRQLVDLHEDVLKEMKAKASTVKALLSTEEACKLTPPESAKSKP 2520
Db 2461 SRSASLRQKVTFRQLVDLHEDVLKEMKAKASTVKALLSTEEACKLTPPESAKSKP 2520
Qy 2521 GYGAKVNRNLSRAVNHRSVWEDLLEDTEPIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
Db 2521 GYGAKVNRNLSRAVNHRSVWEDLLEDTEPIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
Qy 2581 FPDIGVAVCEKMAIDYVSTLPPQAVMGSSYGFQSPQORVEFLVNTWKSCKCPMGFSYDT 2640
Db 2581 FPDIGVAVCEKMAIDYVSTLPPQAVMGSSYGFQSPQORVEFLVNTWKSCKCPMGFSYDT 2640
Qy 2641 RCFDSTVTESDIRVEESIOCCDLAPARQAIRSLTERLYTGGPLTNSKGQNCYRRCRA 2700
Db 2641 RCFDSTVTESDIRVEESIOCCDLAPARQAIRSLTERLYTGGPLTNSKGQNCYRRCRA 2700
Qy 2701 SGVLTTCGNTLTCLYKATAACRAAKLQDCITMLVNGDDLVCISAGTQEDAAALRAFTE 2760
Db 2701 SGVLTTCGNTLTCLYKATAACRAAKLQDCITMLVNGDDLVCISAGTQEDAAALRAFTE 2760
Qy 2761 AMTHYSAPPGDPPOPEVDLELITSCSNVSVAHSDASGRVYLLTRDPTPLARAAMETAR 2820
Db 2761 AMTHYSAPPGDPPOPEVDLELITSCSNVSVAHSDASGRVYLLTRDPTPLARAAMETAR 2820
Qy 2821 HTPINSWLGNIIMYAPTILWARMILMTHFFSLLAQEOLKALDCQIYGACYSIEPLDLPO 2880
Db 2821 HTPINSWLGNIIMYAPTILWARMILMTHFFSLLAQEOLKALDCQIYGACYSIEPLDLPO 2880
Qy 2881 IIERLHGLSAPTHSYSGEINRVASCLRKLGVPLRTWHRARSVRAKLISQGRAATC 2940
Db 2881 IIERLHGLSAPTHSYSGEINRVASCLRKLGVPLRTWHRARSVRAKLISQGRAATC 2940
Qy 2941 GRXILFNWAVRTKLTIPAAQSOLDLGSFWAGYSGGDIYHLSRARPRPFCLILLISV 3000
Db 2941 GRXILFNWAVRTKLTIPAAQSOLDLGSFWAGYSGGDIYHLSRARPRPFCLILLISV 3000
Qy 3001 GVGIIYLLPNR 3010
Db 3001 GVGIIYLLPNR 3010
```

```
RESULT 15
Q9QIY4 PRELIMINARY; PRT; 3010 AA.
ID Q9QIY4
AC Q9QIY4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RC STRAIN=MD4-1;
RC STRAIN=MD4-1;
RC MEDLINE=20013325; PubMed=10544098;
RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.-Y., Miyasaka Y.,
RA Tazawa J.-i., Izumi N., Marumo F., Sato C.;
RT "Time-related changes in full-length hepatitis C virus and hepatitis
RT activity";
RL Virology 263:244-253(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MD4-1;
RA Nagayama K., Kurosaki M., Maekawa S., Maekawa S., Miyasaka Y.,
RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MENA (BY SIMILARITY).
CC EMBL; AF165051; AAD56186.1; -.
DR PIR; A61136; A61196.
DR PIR; P0246; P0246.
DR PIR; P0804; P0804.
DR PIR; P0329; P0329.
DR HSP; P26663; 1JXP.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR01410; DEAD.
DR InterPro; IPR02522; HCV core.
DR InterPro; IPR02521; HCV core.
DR InterPro; IPR02519; HCV env.
DR InterPro; IPR02531; HCV NS1.
DR InterPro; IPR02515; HCV NS2.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR02166; HCV RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_C29.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01539; HCV core; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
```

DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326830 MW; F6559AB2CFC53CB2 CRC64;

Query Match 95.7%; Score 15326; DB 12; Length 3010;
 Best Local Similarity 94.8%; Pred. No. 0;
 Matches 2853; Conservative 79; Mismatches 78; Indels 0; Gaps 0;

QY 1 XSTNPQKTKENTNRPPQDKVFPGGQIVGGVYLLPRGPRGLGVRAATKASERSQPRG 60
 DB 1 XSTNPQKTKENTNRPPQDKVFPGGQIVGGVYLLPRGPRGLGVRAATKTSERSQPRG 60

QY 61 RQPIPKARPEGRWAQPGYPWPLVYNEGLGWAGWMLSPRGSRPQSGPTDPRRRNLG 120
 DB 61 RQPIPKARPEGRWAQPGYPWPLVYNEGLGWAGWMLSPRGSRPQSGPTDPRRRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVAGPLGAARALAHGVRLVDGVNATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPLVAGPLGAARALAHGVRLVDGVNATGNLPGCSFSIFLLA 180

QY 181 LLSCLTIPASAEVRNVSGIHYHVTNDCSNSSIVYEADVIMHTPGVPCVQEGNSRQWV 240
 DB 181 LLSCLTIPASAEVRNVSGIHYHVTNDCSNSSIVYEADVIMHTPGVPCVQEGNSRQWV 240

QY 241 ALPTTLAARNASVPTTIRRHVDLLVGTAAFCASMYVDLCSIFLVSQLFTSPRRHET 300
 DB 241 ALPTTLAARNASVPTTIRRHVDLLVGTAAFCASMYVDLCSIFLVSQLFTSPRRHET 300

QY 301 VQDCNCSIVPGHVSGRMAWMMWSPPTALVVSQLLRIPOAVDVMWAGHGVLAGLA 360
 DB 301 VQDCNCSIVPGHVSGRMAWMMWSPPTALVVSQLLRIPOAVDVMWAGHGVLAGLA 360

QY 361 YYSWGNWAKVILLVALLFAGVDGETHTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420
 DB 361 YYSWGNWAKVILLVALLFAGVDGETHTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420

QY 421 HNRTALNCDSLOTGFFAALFYAHFNSSGCGPERMASCRPIDWFAQGWGPIITYKPNSS 480
 DB 421 HNRTALNCDSLOTGFFAALFYAHFNSSGCGPERMASCRPIDWFAQGWGPIITYKPNSS 480

QY 481 DQRPYCHWYAPRCGVVPASQGVVYCTFSPVWVGTTRDSGPVYTSWGENETDVMLLN 540
 DB 481 DQRPYCHWYAPRCGVVPASQGVVYCTFSPVWVGTTRDSGPVYTSWGENETDVMLLN 540

QY 541 NTRPPQGNWFGCTWMSGTGTCGGPPONIGVGNRTLICPTDCPRKHPEATYTKCGSG 600
 DB 541 NTRPPQGNWFGCTWMSGTGTCGGPPONIGVGNRTLICPTDCPRKHPEATYTKCGSG 600

QY 601 PMLTTPCLVDYDYLRLMHPCTLNFSIFKVRMYGVGVEHRLNAACNTRGRCNLEDRDS 660
 DB 601 PMLTTPCLVDYDYLRLMHPCTLNFSIFKVRMYGVGVEHRLNAACNTRGRCNLEDRDS 660

QY 661 EUSPLLSSTTEWQILPCAFITLTPALSTGLHLHQNIVDVQYLYGVGSFAFVSFAIKWEYIL 720
 DB 661 EUSPLLSSTTEWQILPCAFITLTPALSTGLHLHQNIVDVQYLYGVGSFAFVSFAIKWEYIL 720

QY 721 LLELLADARVCACLWMMLLIAQAEALLENVLVNLNAAVAGAHGILSFLVFFCAAWYIKG 780
 DB 721 LLELLADARVCACLWMMLLIAQAEALLENVLVNLNAAVAGAHGILSFLVFFCAAWYIKG 780

QY 781 RLAPGAAYAFYGVWPIILLALLALPPRAYALDREMAASCGGAVLVGVFTLSPYKVELT 840
 DB 781 KLVPGRAYAIYGVWPIILLALLALPPRAYALDREMAASCGGAVLVGVFTLSPYKVELT 840

QY 841 RLHWLQYFTRAEAHQWVWVPPNVGRGDDAILLTCAVHEPILIFDTIKMLLAILGPLM 900
 DB 841 RLHWLQYFTRAEAHQWVWVPPNVGRGDDAILLTCAVHEPILIFDTIKMLLAILGPLM 900

QY 901 VLOAGITRVPEVRAQGLIRACMLVRKVGAGHYQVFMKLGALGTGVYVNHILTDLRWA 960
 DB 901 VLOASIVRVPEVRAQGLIRACMLVRKKAAGHYAQWVLVRLAALGTGVYVNHILTDLRWA 960

QY 961 HAGLRDLAVAPVVPVPSAMETKVTITWAGDTAACGDIIILGLPVSAARGKEIFLGPADSLEG 1020
 DB 961 HAGLRDLAVAPVVPVPSAMETKVTITWAGDTAACGDIIISGLPVSAARGKEILGPADSLEG 1020

QY 1021 QQWRLAPITAYSQOTRGVLGCIITSLTGRDKNQVEGEVQVVTATQSFATPCINGCVWT 1080
 DB 1021 QQWRLAPITAYSQOTRGVLGCIITSLTGRDKNQVEGEVQVVTATQSFATPCINGCVWT 1080

QY 1081 VYHAGSKXTLAGPKGPIITQMTYTNVDLIVGQAPPGARSMTFPCSSGSDLYLVTRHADVI 1140
 DB 1081 VYHAGSKXTLAGPKGPIITQMTYTNVDLIVGQAPPGARSMTFPCSSGSDLYLVTRHADVI 1140

QY 1141 PYRRGRDGRGSLSPRPVSYLKGSGGGLLCPSHGVVGVFRAAIVCTRGVAKAVDIPVES 1200
 DB 1141 PYRRGRDGRGSLSPRPVSYLKGSGGGLLCPSHGVVGVFRAAIVCTRGVAKAVDIPVEA 1200

QY 1201 METTMRSPVFTDNSTPPPAVPOTFQVAHLHAPTGSGKSTKVPAAVAAQGVKVLVNPVAA 1260
 DB 1201 METTMRSPVFTDNSTPPPAVPOTFQVAHLHAPTGSGKSTKVPAAVAAQGVKVLVNPVAA 1260

QY 1261 TLGFGAYMSKAHGDINRTGVRTITTTGGSITYSTYKFLADGGSGGAYDIIICDECHS 1320
 DB 1261 TLGFGAYMSKAHGDINRTGVRTITTTGGSITYSTYKFLADGGSGGAYDIIICDECHS 1320

QY 1321 TDSTWILGTVLDOAETAGARLVLATATPGSVTVVPHNTEELCSNNGEIPYKAI 1380
 DB 1321 TDSTWILGTVLDOAETAGARLVLATATPGSVTVVPHNTEELCSNNGEIPYKAI 1380

QY 1381 PTEALKGGEHLIFCHSKKCDLAAKLTGLGNNAVYVYRGLDVSVIPPIGDVVVATDAL 1440
 DB 1381 PTEALKGGEHLIFCHSKKCDLAAKLTGLGNNAVYVYRGLDVSVIPPIGDVVVATDAL 1440

QY 1441 MTGFTGDFSDVIDNCTVQTVDPSLDTFTTETTTVPQDAVSRQRGRTRGRSGIYR 1500
 DB 1441 MTGFTGDFSDVIDNCTVQTVDPSLDTFTTETTTVPQDAVSRQRGRTRGRSGIYR 1500

QY 1501 FYTPGERPSGMDSSVLCECYDAGCAWVELTFAETSVRLRAYLNTPLPVCQDHLFEFWS 1560
 DB 1501 FYTPGERPSGMDSSVLCECYDAGCAWVELTFAETSVRLRAYLNTPLPVCQDHLFEFWS 1560

QY 1561 VFTGTLTHIDAHFSLSTKOAGDNFPYLVAYQATVCAARAQPPSDQMKCLIRLXPTLHG 1620
 DB 1561 VFTGTLTHIDAHFSLSTKOAGDNFPYLVAYQATVCAARAQPPSDQMKCLIRLXPTLHG 1620

QY 1621 PTPLLYRIGAVONEVILTHPIIKYINACMSADLEVVTSTWLVGGVLAALAAAYCLTTGSV 1680
 DB 1621 PTPLLYRIGAVONEVILTHPIIKYINACMSADLEVVTSTWLVGGVLAALAAAYCLTTGSV 1680

QY 1681 VTVGRILISGKPAVDPDRVLYQEFDEMEECASQIPYIEQGMQLAEQFKQKALGLLOTAT 1740
 DB 1681 VTVGRILISGKPAVDPDRVLYQEFDEMEECASQIPYIEQGMQLAEQFKQKALGLLOTAT 1740

QY 1741 KQAEAAAPVSVESKWRALETFNAKEMWNFTISGQYLAGLSTLPGNPAIASLMAFTASITSP 1800
 DB 1741 KQAEAAAPVSVESKWRALETFNAKEMWNFTISGQYLAGLSTLPGNPAIASLMAFTASITSP 1800

QY 1801 LTTQNTLIFNLIGWVAAQLAPPSAASAFVVGAGTAAVGSIGLQKLVLDILAGYGAGVA 1860
 DB 1801 LTTQNTLIFNLIGWVAAQLAPPSAASAFVVGAGTAAVGSIGLQKLVLDILAGYGAGVA 1860

QY 1861 GALVAFKMSGVPSTEDLVNLLPAILSPGALVGVCAAILRRHVGPGEVQVWNNRLI 1920
 DB 1861 GALVAFKMSGVPSTEDLVNLLPAILSPGALVGVCAAILRRHVGPGEVQVWNNRLI 1920

QY 1921 AFASRGNHVSPHYVPESDAARVTQIILSSLTITQLKRLHOWINEDCSTPCSGSWLRDV 1980
 DB 1921 AFASRGNHVSPHYVPESDAARVTQIILSSLTITQLKRLHOWINEDCSTPCSGSWLRDV 1980

Search completed: April 8, 2004, 10:48:58
Job time : 105 secs

QY	1981	WDWICTVLTDKFKTQSKLPRLPVFPFLSCQRYKGVWRGDMQTTCPCCAGIAGVK	2040
Db	1981	WDWICTVLSDFKTWLQSKLPRLPGLFPLSCQRYKGVWRGDMQTTCPGAGITGEVK	2040
QY	2041	NGSMRIIVGPRTCSTNWTGTPPINAYTTGCPDPSAPNYSRALWRAAEYVEVTRVGDFH	2100
Db	2041	NGSMRIIVGPRTCSTNWTGTPPINAYTTGCPDPSAPNYSRALWRAAEYVEVTRVGDFH	2100
QY	2101	YVTCMTTNDNVKPCQVPAPEPTEVDGVRHRYAPACKPLRLREDDVTQVGLNQYLVGSOL	2160
Db	2101	YVTCMTTNDNVKPCQVPAPEPTEVDGVRHRYAPACKPLRLREDDVTQVGLNQYLVGSOL	2160
QY	2161	PCEPEPDVTVLTSMLTDPSHITAKERLARLARGSPPLASSASQLSAPSLKATCTTHHD	2220
Db	2161	PCEPEPDVAVLTSMLTDPSHITAKERLARLARGSPPLASSASQLSAPSLKATCTTHHD	2220
QY	2221	SPDADLIEANLLWQEMGNNITRVESNKVVILDSFPLHAGDERBISVAAEILRKSRK	2280
Db	2221	SPDADLIEANLLWQEMGNNITRVESNKVVILDSFPLHAGDERBISVAAEILRKSRK	2280
QY	2281	FPSALPIWARPDYNPPLLESWKDPDYPPVVGCPPLPTKAPPIPPRRKRTVVLTSNV	2340
Db	2281	FPMAMPVWARPDYNPPLLESWKDPDYPPVVGCPPLPTKAPPIPPRRKRTVVLTSNV	2340
QY	2341	SSALAEATKTFGSSGSAVDSGTATAPDLASDDGDKGSDVESYSMPPLGEGPDPDL	2400
Db	2341	SSALAEATKTFGSSGSAVDSGTATAPDLASDDGDKGSDVESYSMPPLGEGPDPDL	2400
QY	2401	SDGWSVTSSEASDVVCCSNSTWTGALITPCAABESKLDINPLSNLRRHNNVYATT	2460
Db	2401	SDGWSVTSSEASDVVCCSNSTWTGALITPCAABESKLDINPLSNLRRHNNVYATT	2460
QY	2461	SRSASLRQKVTFRLOVLDHYRDVLKEMKAKASTVKAKLLSTEEACKLTPPHSAKSF	2520
Db	2461	SRSASLRQKVTFRLOVLDHYRDVLKEMKAKASTVKAKLLSTEEACKLTPPHSAKSF	2520
QY	2521	GYGAKDVNLTSSRAVNHRSVWEDLLETPTIDTTIMAKSEVFCVQPEKGRKPARLIY	2580
Db	2521	GYGAKDVNLTGKAVNHRSVWEDLLETPTIDTTIMAKSEVFCVQPEKGRKPARLIY	2580
QY	2581	FPDLGVVCEKMAIYDVVSTLPQAVMGSSYGFQSPKQRFVFLVNTWKSXKCPMGFSYDT	2640
Db	2581	FPDLGVVCEKMAIYDVVSTLPQAVMGSSYGFQSPKQRFVFLVNTWKSXKCPMGFSYDT	2640
QY	2641	RCFDSVTESDIRVEESIYQCCDLAPARQAIKSLTERLYIGGFLTNSKGQNCYRRCRA	2700
Db	2641	RCFDSVTENDIRVEESIYQCCDLAPARQAIKSLTERLYIGGFLTNSKGQNCYRRCRA	2700
QY	2701	SGVLTSCGNTLTCYLKATAACRAAKLQDCTMLVNGDDLVCESAGTQEDAAALRAFTE	2760
Db	2701	SGVLTSCGNTLTCYLKASAAKRAAKLQDCTMLVNGDDLVCESAGTQEDAAALRAFTE	2760
QY	2761	AMTHYSAPGPPQPEYDLELITSCSNVSVAHASGKRVYLTEDPTTPLARAWEETAR	2820
Db	2761	AMTHYSAPGPPQPEYDLELITSCSNVSVAHASGKRVYLTEDPTTPLARAWEETAR	2820
QY	2821	HTPINSWLGNIIMVAPTLWARMILMTFFSILLAOEOLKALDCQIYGCYSIBPLDLPO	2880
Db	2821	HTPVNSWLGNIIMVAPTLWARMVLTMTFFSILLAOEOLKALDCQIYGAHYSIEPLDLPO	2880
QY	2881	IIERLHGLSAPFSLHSYSGEINRVASCLRLGVPPFLTRHRAASVRAXLLSQGRAATC	2940
Db	2881	IIERLHGLSAPFSLHSYSGEINRVAACLRLGVPPFLRVHRAASVRAXLLSQGRAATC	2940
QY	2941	GSYLFNMAVTRKILTPIPASQDLDSWFVAGYSGGDIYHLSRARPWFPLCLLLSV	3000
Db	2941	GSYLFNMAVTRKILTPIPASQDLDSWFVAGYSGGDIYHLSRARPWFPLCLLLSV	3000
QY	3001	GVGIYLLPNR 3010	
Db	3001	GVGIYLLPNR 3010	

DR PIR; A32553; GNVWCJ.
 DR HSS2; E26683; LUXP.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV env.
 DR InterPro; IPR002531; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002518; HCV NS1.
 DR InterPro; IPR002517; HCV NS2.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002668; HCV NS2a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR004109; Peptidase C29.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol PSvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SMC0487; DRXDe; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match

96.0%; Score 15367; DB 1; Length 3010;

Best Local Similarity 95.2%; Pred. No. 0;
 Matches 2866; Conservative 64; Mismatches 80; Indels 0; Gaps 0;
 QY 1 MSTNPKPQKTKRNTNRPPQDVKPGGGQIYGVYLLPRRGRPRGVRATRKASRSQPRG 60
 DB 1 MSTNPKPQKTKRNTNRPPQDVKPGGGQIYGVYLLPRRGRPRGVRATRKASRSQPRG 60
 QY 61 RROPIKARRPEGRRAWAOPGYFWPLYGNEGLWAGWLLSPRGRSPSWGTFDPRRSRLG 120
 DB 61 RROPIKARRPEGRRAWAOPGYFWPLYGNEGLWAGWLLSPRGRSPSWGTFDPRRSRLG 120
 QY 121 KVIDTLCTGADLMGYIPLVGAPLGGARALAHGVRLVDGNYATGTLNLPCCSIFILIA 180
 DB 121 KVIDTLCTGADLMGYIPLVGAPLGGARALAHGVRLVDGNYATGTLNLPCCSIFILIA 180
 QY 181 LLSCLTIPASAYEVNRVSGIYHVNDCSNISIYVEADVIHMTPCGVCVCEGNSRCWV 240
 DB 181 LLSCLTIPASAYEVNRVSGIYHVNDCSNISIYVEADVIHMTPCGVCVCEGNSRCWV 240
 QY 241 ALTPTLAARNASVPTTIRRHVDLLAVGTAAPCSAMVVDLCSIFLVLSQTFPFRHET 300
 DB 241 ALTPTLAARNASVPTTIRRHVDLLAVGTAAPCSAMVVDLCSIFLVLSQTFPFRHET 300
 QY 301 VQDCNCSIYPGHVSHERVAMMMWSPFTALVVSQLLRIPOAVVDMVAGAHGVLGLA 360
 DB 301 VQDCNCSIYPGHVSHERVAMMMWSPFTALVVSQLLRIPOAVVDMVAGAHGVLGLA 360
 QY 361 YYSWGVNNAKVLIVALLFAGVDGTHVTVGRVAGHTTSGTSLFFSSGASQKQLVNTGWS 420
 DB 361 YYSWGVNNAKVLIVALLFAGVDGTHVTVGRVAGHTTSGTSLFFSSGASQKQLVNTGWS 420
 QY 421 HINRTALNCNDSLOTGFAALFYAHKFNSSGCPERVASCRIPIFAQKGPITYTKPSS 480
 DB 421 HINRTALNCNDSLOTGFAALFYAHKFNSSGCPERVASCRIPIFAQKGPITYTKPSS 480
 QY 481 DORPYCHVAPRPGCVVPASQVCGPVYCTPSPVVTGTDGSGVTYSWGENETDVLIN 540
 DB 481 DORPYCHVAPRPGCVVPASQVCGPVYCTPSPVVTGTDGSGVTYSWGENETDVLIN 540
 QY 541 NTRPQGNWFGCTWNSTGFTKTCGGPPCNI GGVNRTLICPTDCKFRKPEATYTKCSG 600
 DB 541 NTRPQGNWFGCTWNSTGFTKTCGGPPCNI GGVNRTLICPTDCKFRKPEATYTKCSG 600
 QY 601 PMLTRCLVDYPRLWHVYPTLNFISIFKVMVVGVEHRLNACNWTGERCNLEDRDS 660
 DB 601 PMLTRCLVDYPRLWHVYPTLNFISIFKVMVVGVEHRLNACNWTGERCNLEDRDS 660
 QY 661 ELSPLLSTTEWQILPCARTLPAI STGLIHHLHONIVDVQYLYGVGSAPVSAIKWEYIL 720
 DB 661 ELSPLLSTTEWQILPCARTLPAI STGLIHHLHONIVDVQYLYGVGSAPVSAIKWEYIL 720
 QY 721 LFLLLADARVCACLMMWLLIAQAEEAENLVVLAASVAGAGHLLSFLVFFCAAWYIKG 780
 DB 721 LFLLLADARVCACLMMWLLIAQAEEAENLVVLAASVAGAGHLLSFLVFFCAAWYIKG 780
 QY 781 RLAPGAAYAFYGVWPLLLILLALIPRAYALDEMAASCGGAVLVGVFTLSPYKVELT 840
 DB 781 RLAPGAAYAFYGVWPLLLILLALIPRAYALDEMAASCGGAVLVGVFTLSPYKVELT 840
 QY 841 RLHWMLOYFTTRAEAHMVVPPNVRGGRDAIILLTCAVHPDLIPDITKLLAILGLPLM 900
 DB 841 RLHWMLOYFTTRAEAHMVVPPNVRGGRDAIILLTCAVHPDLIPDITKLLAILGLPLM 900
 QY 901 VLQAGITRVPYFVRAOGLIRACMLVKVAGGHVQVMFKLGALGTYYYNHLLTPRLWA 960
 DB 901 VLQAGITRVPYFVRAOGLIRACMLVKVAGGHVQVMFKLGALGTYYYNHLLTPRLWA 960
 QY 961 HAGRLDAVABEVVVSAMETKVIITWGDATACCDIILGLPVSARKKILFGLPADSLG 1020
 DB 961 HAGRLDAVABEVVVSAMETKVIITWGDATACCDIILGLPVSARKKILFGLPADSLG 1020
 QY 1021 QGWELLAPITAYSQOTRGVLGCIITITGDRKNQVEGEVQVWSTATOSFLATCINGVCWT 1080

Db 1021 QGWRLLAPITAYSQQRGLLGCITSLTGRDKQVDCGEVOVLSTATQSFPLATCVNGVCWT 1080
QY 1081 VYHGAGSKTLAGPKGPTOMYTNVLDLVCWQAPPGARSMTPCSCSSDLYLYTRHADYI 1140
Db 1081 VYHGAGSKTLAGPKGPTOMYTNVQDLQWGPAPPGARSMTPTCCSSDLYLYTRHADYV 1140
QY 1141 PVRRRGDSRGLSLSPRPVSYLKSSGGPPLCPSGHVGVFRAAVCTRGVAKAVDFIPVES 1200
Db 1141 PVRRRGDSRGLSLSPRPISYLKSSGGPPLCPSGHVGVIFRAAVCTRGVAKAVDFIPVES 1200
QY 1201 METTMRSPVTDNSTPFAVPTQFOVAHLHAPTSGKSTKVPAAVAAQYKVLVLPNSVAA 1260
Db 1201 METTMRSEVTDNSTSPPAVPTQFOVAHLHAPTSGKSTKVPAAVAAQYKVLVLPNSVAA 1260
QY 1261 TLGFGAYSKAHGTDNIRGVRTITGGSTVSTYKELADGSCSGAYDIIICDRCHS 1320
Db 1261 TLGFGAYSKAHGIEPNIRGVRTITGGPIITSTYCKFLADGSCSGAYDIIICDRCHS 1320
QY 1321 TDSTTILIGITVLDQAETAGARLVVLATATPPGSVTVPHENIEIGLSNNGEIPFYGKAI 1380
Db 1321 TDSTTILIGITVLDQAETAGARLVVLATATPPGSIITVPHENIEEVALSNTGEIPFYGKAI 1380
QY 1381 PIRAIKGRHLIFCHSKKKDELAALKLTGLNAYAYRGJDSVVIPIGDVVVVATDAL 1440
Db 1381 PIRAIKGRHLIFCHSKKKDELAALKLTGLNAYAYRGJDSVVIPTSGDVVVVATDAL 1440
QY 1441 MTGFTGDFDSVIDNCVTOTVDPSLDPTFTIETTVPDQAVSRORSRGRGRSGSIYR 1500
Db 1441 MTGFTGDFDSVIDNCVTOTVDPSLDPTFTIETTVPDQAVSRORSRGRGRSGSIYR 1500
QY 1501 FVTPGERPSGMFDSVILCECYDAGCAWYELTFAETSVRLRAYLNTPLGVCQDHLFEWES 1560
Db 1501 FVTPGERPSGMFDSVILCECYDAGCAWYELTFAETSVRLRAYLNTPLGVCQDHLFEWES 1560
QY 1561 VFTGLTHIDAHFLSQTQWAGNFPYLVAQYATVCARAQAAPPSPWDQWKKLIRLKPILHG 1620
Db 1561 VFTGLTHIDAHFLSQTQWAGNFPYLVAQYATVCARAQAAPPSPWDQWKKLIRLKPILHG 1620
QY 1621 PTPLLYRLGAVQNSVILTHPTTKYIMACMSADLEVTSTWLVGVGLAALAAAYCLTTGSV 1680
Db 1621 PTPLLYRLGAVQNSVILTHPTTKYIMACMSADLEVTSTWLVGVGLAALAAAYCLTTGSV 1680
QY 1681 VIVGRIITLSGKPAYVPDREVLVYQFDMEECASQJPIYIEQGMQLAEQFKQKALGLLOTAT 1740
Db 1681 VIVGRIITLSGKPAVVPDREVLVYQFDMEECASLXPVIEQGMQLAEQFKQKALGLLOTAT 1740
QY 1741 KOABAAAPVVEKWRALLETFAKHMWNPISGIOHLAGLSTLFGNPAIASLMAFTASTSP 1800
Db 1741 KOABAAAPVVEKWRALLETFAKHMWNPISGIOHLAGLSTLFGNPAIASLMAFTASTSP 1800
QY 1801 LTTQNTLLFNILGWVAAQLAPPAAAFVAGIAGAAVGSIGLGKVLVDILAGYGAGVA 1860
Db 1801 LTTQNTLLFNILGWVAAQLAPPAAAFVAGIAGAAVGSIGLGKVLVDILAGYGAGVA 1860
QY 1861 GALVAFKVMSCGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGEVAVOMNRLLI 1920
Db 1861 GALVAFKVMSCGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGEVAVOMNRLLI 1920
QY 1921 AFASRGHVSPTHVPSSDAARVTOILSSITITOLLKRLHOWNEDCSTPCSSGSLRDV 1980
Db 1921 AFASRGHVSPTHVPSSDAARVTOILSSITITOLLKRLHOWNEDCSTPCSSGSLRDV 1980
QY 1981 WDMICTVLTDKTLQSKLLPRLPGVPFELSCORGKGVWRGDGMQITPCGQAIAGHVK 2040
Db 1981 WDMICTVLSDKTKLQSKLLPRLPGLPFLSQRGKGVWRGDGLVQITPCGQAITGVK 2040
QY 2041 NGSXRIUGPRTCSNTHGTFPPINAYTTGCTPSPAPNYSRALWRVAABEYVEVTRVGDFH 2100
Db 2041 NGSXRIUGPKTCSNTHGTFPPINAYTTGCTPSPAPNYSRALWRVAABEYVEVTRVGDFH 2100
QY 2101 YVTGMTDNDVKPCQVPAPEPTEVDGVLRLHRYAPACKPLLRDYTFVQGLNQKLVGSQL 2160
Db 2101 YVTGMTDNDVKPCQVPAPEPTEVDGVLRLHRYAPACKPLLRDYTFVQGLNQKLVGSQL 2160

QY 2161 PCEPEPDVTLTSMITDPSHITAEAKRLARGSPPSLASSASQSLASPSLKATCTTHHD 2220
Db 2161 PCEPEPDVAVLTXLITDPSHITAEAKRLARGSPPSLASSASQSLASPSLKATCTTHHD 2220
QY 2221 SPDADLIEANLLWRQMGNGNITRVESENKVVILDSPEPLHAEGDREIRISVAABILKSKR 2280
Db 2221 SPDADLIEANLLWRQMGNGNITRVESENKVVILDSFPDIRAVEDEIRISVPAABILKPKR 2280
QY 2281 FPSALFIWARPDPNPLLESWKDPDVPVWVGCPPLPPTKAPPIPPRKRRTVVLTFESV 2340
Db 2281 FPALEFIWARPDPNPLLESWKDPDVPVWVGCPPLPPTKAPPIPPRKRRTVVLTFESTV 2340
QY 2341 SSALAEALATKTFGSSGSSAVDSGTATATPDLASDDGKSGDSVESYSGMPPLEGEPPDPL 2400
Db 2341 SSALAEALATKTFGSSGSSAVDSGTATATPDLASDDGKSGDSVESYSGMPPLEGEPPDPL 2400
QY 2401 SDGWSVTVEASEEDVVCSSMSYWTGALITPCAABESKLPINPLNSLIRHHNMVYATT 2460
Db 2401 SDGWSVTVEASEEDVVCSSMSYWTGALITPCAABESKLPINPLNSLIRHHNMVYATT 2460
QY 2461 SRASLRQKVTDFRLQVLDHRYDLKEMKAKASTVKAKLISIEACKLTPEHSAKSKF 2520
Db 2461 SRASLRQKVTDFRLQVLDHRYDLKEMKAKASTVKAKLISIEACKLTPEHSAKSKF 2520
QY 2521 GYGAKDVRNLSRAVNHIRSVWEDLLEDTETPIDTTIMAKSEVFCVQPEKGGKPKARLIV 2580
Db 2521 GYGAKDVRNLSRAVNHIRSVWEDLLEDTETPIDTTIMAKSEVFCVQPEKGGKPKARLIV 2580
QY 2581 FPDGLVRVCEKMALYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTWKSCKCPMGFSYDT 2640
Db 2581 FPDGLVRVCEKMALYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTWKSCKCPMGFSYDT 2640
QY 2641 RCFDSTVTESDIIVESSIYOCCLDAPRAQAIRSLTERLYIGGPLNLSKQNGCYRRCRA 2700
Db 2641 RCFDSTVTENDIIVESSIYOCCLDAPRAQAIRSLTERLYIGGPLNLSKQNGCYRRCRA 2700
QY 2701 SGVLTTCGNTLTCYLKATAACRAAKLQDCTMLVNGDILVVICESAGTQDAAALRAFT 2760
Db 2701 SGVLTTCGNTLTCYLKATAACRAAKLQDCTMLVNGDILVVICESAGTQDAAALRAFT 2760
QY 2761 AMTRYAPPDPPQPPRYDIELITCSNNSVSAHDAAGKRVVYLTTRDPTTPLAANAETAR 2820
Db 2761 AMTRYAPPDPPQPPRYDIELITCSNNSVSAHDAAGKRVVYLTTRDPTTPLAANAETAR 2820
QY 2821 HTPNLSMGLNIMVAPTLMAWMLMTHFPSILLAQOLEKALDCQYGACYSIEPLDLPO 2880
Db 2821 HTPNLSMGLNIMVAPTLMAWMLMTHFPSILLAQOLEKALDCQYGACYSIEPLDLPO 2880
QY 2881 IIERLHGLSAFTLHXSYPGEINRVASCLRLKGVPPPLTWRHARSVRKALLSOGGRAATC 2940
Db 2881 IIERLHGLSAFTLHXSYPGEINRVASCLRLKGVPPPLTWRHARSVRKALLSOGGRAATC 2940
QY 2941 GRVLFNVAVTKLPTIPAASQIDLSGWFVAGYSGDIIYHLSRARPRFPPLCLLLSV 3000
Db 2941 GKYLFNVAVTKLPTIPAASQIDLSGWFVAGYSGDIIYHLSRARPRFPPLCLLLSV 3000
QY 3001 GVGIYLLPNR 3010
Db 3001 GVGIYLLPNR 3010

RESULT 2
POLG HCVEK STANDARD; PRT; 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (p22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (p21)
DE (EC 3.4.22.-); Protease/helicase NS3 (p70) (Hepacivirin)

(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
Hepatitis C virus (isolate BK) (HCV).
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
NCBI_taxonomy=11105;
[1]
SEQUENCE FROM N.A.
MEDLINE=911140698; PubMed=1847440;
Takamiwawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J., Onishi E., Andoh T., Yoshida I., Okayama H.;
"Structure and organization of the hepatitis C virus genome isolated from human carriers.";
J. Virol. 65:1105-1113 (1991).
[2]
SEQUENCE OF 1487-1500.
MEDLINE=96235224; PubMed=8647104;
Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
"Non-structural protein 3 of hepatitis C virus inhibits phosphorylation mediated by cAMP-dependent protein kinase.";
Eur. J. Biochem. 237:611-618 (1996).
[3]
X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
MEDLINE=97015088; PubMed=8861916;
Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N., Moomaw E.W., Adachi T., Hostomska Z.;
"The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site.";
Cell 87:331-342 (1996).
[4]
X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
MEDLINE=98227846; PubMed=9568891;
Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M., Steinkuehler C., Tomai L., de Francesco R., Kuo L.C., Chen Z.;
"Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2-A resolution structure in a hexagonal crystal form.";
Protein Sci. 7:837-847 (1998).
-!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.
-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
-!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.
-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcement/> or send an email to license@ebi.ac.uk).

EMBL; M58335; AAA72945.1; --
PIR; A38465; GNNVTC.
PDB; 1AJO; 25-MAR-98.
PDB; 1JXP; 14-JAN-98.
PDB; 1NS3; 08-APR-98.
PDB; 1C2P; 15-NOV-00.
PDB; 1CSJ; 08-NOV-99.
PDB; 1GX5; 09-APR-02.
PDB; 1GX6; 10-APR-02.
PDB; 1QUV; 26-JUN-00.
PDB; 8OHM; 20-APR-99.
PDB; MEROPS; S29.001; --
PDB; MEROPS; U39.001; --

[illegible]

1681 VIVGRILLISGRPAIVDPRELLIQEEDMEBECAASHLPVIEQWQLAEQFKQKALGLLQAT 1740
1741 KOAEAAAPVESKWALETFFMAKMNPIGQYLAGLSTLPGNPAIASLMAFTASITSP 1800
1741 KOAEAAAPVESKWALETFFMAKMNPIGQYLAGLSTLPGNPAIASLMAFTASITSP 1800
1801 LTTQNTLLFNIJGGWVAAQAPSAASAFVAGIAGAAGVSGISGLKVLVDILLAGVAGVA 1860
1801 LTTQNTLLFNIJGGWVAAQAPSAASAFVAGIAGAAGVSGISGLKVLVDILLAGVAGVA 1860
1861 GALVAFKWSGVEPSTEDJVNLLPAILSPGALVWGVVCAAILRRHVPOGEGAVQWNRLLI 1920
1861 GALVAFKWSGVEPSTEDJVNLLPAILSPGALVWGVVCAAILRRHVPOGEGAVQWNRLLI 1920
1921 AFASRGNHVSPTHYVPESDAARVTOILSSLTITQLLKELHOWINEDCSTPCSGSWLRDV 1980
1921 AFASRGNHVSPTHYVPESDAARVTOILSSLTITQLLKELHOWINEDCSTPCSGSWLRDV 1980
1981 WDWICTVLDFKWLQSKLLPRLPGVFFLSCRGYKGVWGRGIMQITPCGAQIAGHYK 2040
1981 WDWICTVLDFKWLQSKLLPRLPGVFFLSCRGYKGVWGRGIMQITPCGAQIAGHYK 2040
2041 NGSMRIVGPKTCSNTWHTGTFINAYTTGCTSPAPNYSRALWRVAABEYVEVTRVGDHF 2100
2041 NGSMRIVGPKTCSNTWHTGTFINAYTTGCTSPAPNYSRALWRVAABEYVEVTRVGDHF 2100
2101 VYTGMTDNNVKPCQVPAPEFFTEVDGVRHRYAPACKPELLREDVTFQGLNQYLVSQ 2160
2101 VYTGMTDNNVKPCQVPAPEFFTEVDGVRHRYAPACKPELLREDVTFQGLNQYLVSQ 2160
2161 POEPPDVVLTSMLTDPDSHIETAKRRIARLSPSIASSASQLSAPSLKATCTTHD 2220
2161 POEPPDVVLTSMLTDPDSHIETAKRRIARLSPSIASSASQLSAPSLKATCTTHD 2220
2221 SPADUETIANLWQRMGNITRVESKENVILDSPEPLFAEGDEREISVAAILKRSK 2280
2221 SPADUETIANLWQRMGNITRVESKENVILDSPEPLFAEGDEREISVAAILKRSK 2280
2281 FFSALPIWARPDPYNPLLESWKDPDPVPPVHVHGCPLPPTKAPPDPPRKRRTVLTESNV 2340
2281 FFSALPIWARPDPYNPLLESWKDPDPVPPVHVHGCPLPPTKAPPDPPRKRRTVLTESNV 2340
2341 SSALABELATKTFSGSSSAVDSGTATALPDASDDGDKGSDVESYSSMPLEGEPPDPL 2400
2341 SSALABELATKTFSGSSSAVDSGTATALPDASDDGDKGSDVESYSSMPLEGEPPDPL 2400
2401 SDGSWSTVSEAEEDVVCSSMYTWTGALITPCAABESKLPINPLNSLRHHNNVYAT 2460
2401 SDGSWSTVSEAEEDVVCSSMYTWTGALITPCAABESKLPINPLNSLRHHNNVYAT 2460
2461 SRSASLRQKVTDPRLQVLDDHRYDLKEMKAKASTVAKKLSVEEACKLTPPHSAKSF 2520
2461 SRSASLRQKVTDPRLQVLDDHRYDLKEMKAKASTVAKKLSVEEACKLTPPHSAKSF 2520
2521 GYKAKDVRLSSAVKHSWEDLEDETPTDITIMAKSEVFCVQPEKGGKPARLIV 2580
2521 GYKAKDVRLSSAVKHSWEDLEDETPTDITIMAKSEVFCVQPEKGGKPARLIV 2580
2581 FPDGLVRVCEKVALXDVSTLPAQVNGSSYGFQSPKQRFVFLVNTWKSCKCPMGFSYDT 2640
2581 FPDGLVRVCEKVALXDVSTLPAQVNGSSYGFQSPKQRFVFLVNTWKSCKCPMGFSYDT 2640
2641 RCFDSTVTESDIRVESIYOCCLAPAROAISLTERLYIGGLNKSQNGCYRRCRA 2700
2641 RCFDSTVTESDIRVESIYOCCLAPAROAISLTERLYIGGLNKSQNGCYRRCRA 2700
2701 SGVLTWSCGNTLTCYLKATACRAKLODCTMLVNGDDLVIICESAGTOEDAAALFAFE 2760
2701 SGVLTWSCGNTLTCYLKATACRAKLODCTMLVNGDDLVIICESAGTOEDAAALFAFE 2760
2761 AMTRYAPPDPDPQPEYDLELITSCSSNVSAVHADSGKRVYVLTTRDPTTFLAANAWEAR 2820
2761 AMTRYAPPDPDPQPEYDLELITSCSSNVSAVHADSGKRVYVLTTRDPTTFLAANAWEAR 2820

2821 HTPINSWLNGLIIMVAPTILWARMILMTHTFFSILLAQEQLKALDQIYGACYSIEPLDLPQ 2880
2821 HTPINSWLNGLIIMVAPTILWARMILMTHTFFSILLAQEQLKALDQIYGACYSIEPLDLPQ 2880
2881 IIEELHGLSAPTILHSYFGEINRVASCLKLVGPPLTWRHARSVEAKLLSQGGRATC 2940
2881 IIEELHGLSAPTILHSYFGEINRVASCLKLVGPPLTWRHARSVEAKLLSQGGRATC 2940
2941 GRYLFNNAVTRKLTPTPAASQDLSCWFWAGVSGGDIYHSLSRARPRWFPPLCLLLISV 3000
2941 GRYLFNNAVTRKLTPTPAASQDLSCWFWAGVSGGDIYHSLSRARPRWFPPLCLLLISV 3000
3001 GVGIVYLLPNR 3010
3001 GVGIVYLLPNR 3010

RESULT 3
POLG HCVJT
ID POLG HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]_TaxID=31642;
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals."
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D11153; BAA01943.1; -.
CC FIC; A45573; A45573.
CC MEROPS; S29.001; -.
CC InterPro; IPR009003; Cys_ser_trypsin.
CC InterPro; IPR001410; DEAD.
CC DR

Db 1081 VTHGAGSKTLAGPKGPIQMTYNNVDQLVGVHAPPGARSLTPTCTCGSDLYLVREADVI 1140
QY 1141 PVRRGDSRGSLSPRVSVYLKSSGGPLCPSPGHVGVFRFAAVCTRGVAKAVDFIPVES 1200
Db 1141 PVRRGDSRGSLSPRVSVYLKSSGGPLCPSPGHVGVFRFAAVCTRGVAKAVDFIPVES 1200
QY 1201 METMRSPVFTDSTPPAVPQTFOVAHLHAPTSGKSTKVPAAVAAQYKVLVNPSSVAA 1260
Db 1201 METMRSPVFTDSTPPAVPQTFOVAHLHAPTSGKSTKVPAAVAAQYKVLVNPSSVAA 1260
QY 1261 TLGAGVMSXAHGIDPNIRGVRITITGSGTITSTYTKFADGCGSGAYDIICDECHS 1320
Db 1261 TLGAGVMSXAHGIDPNIRGVRITITGSGTITSTYTKFADGCGSGAYDIICDECHS 1320
QY 1321 TDSITILGIGTLDQAETAGARLVVLTATPPGSVTVPHNIEBIGLSNNGEIPFYGKAI 1380
Db 1321 TDSITILGIGTLDQAETAGARLVVLTATPPGSVTVPHNIEBIGLSNNGEIPFYGKAI 1380
QY 1381 PIERAKGRHLIFCHSKKKDELAAKLTGLGNVAVYRGLDVSVIPIGDVVVATDAL 1440
Db 1381 PIERAKGRHLIFCHSKKKDELAAKLTGLGNVAVYRGLDVSVIPIGDVVVATDAL 1440
QY 1441 MTGFTGPFDSVIDCNTCVTQVDFSLDPTFTTITTPQDAVSRQRGRGSGIYR 1500
Db 1441 MTGFTGPFDSVIDCNTCVTQVDFSLDPTFTTITTPQDAVSRQRGRGSGIYR 1500
QY 1501 FVTPGERPSGMDSSVLCYCEYDAGCAWYELTPAETSVRLRAYLNTPLGVCQDHFLES 1560
Db 1501 FVTPGERPSGMDSSVLCYCEYDAGCAWYELTPAETSVRLRAYLNTPLGVCQDHFLES 1560
QY 1561 VFTGHTHIDAHFLSQTQAGNPFYLVAYQATVCARAQAPPSQDQWKLIRLKPILHG 1620
Db 1561 VFTGHTHIDAHFLSQTQAGNPFYLVAYQATVCARAQAPPSQDQWKLIRLKPILHG 1620
QY 1621 PTPLLYRLGAVQNEVITHTPTTKYIMACMSADLEWVTSVWLVGVLAALAAAYCLITGVS 1680
Db 1621 PTPLLYRLGAVQNEVITHTPTTKYIMACMSADLEWVTSVWLVGVLAALAAAYCLITGVS 1680
QY 1681 VIVGRIILSGKPAVVPREVLYOBFDEMERCASQLPYIEQGMQLAOFKQKALGLLOTAT 1740
Db 1681 VIVGRIILSGKPAVVPREVLYOBFDEMERCASQLPYIEQGMQLAOFKQKALGLLOTAT 1740
QY 1741 KQRAAAPVYESKWRALFTWAKHMWFIQIYVLAGSLTLPGNPATASMAFTASITSP 1800
Db 1741 KQRAAAPVYESKWRALFTWAKHMWFIQIYVLAGSLTLPGNPATASMAFTASITSP 1800
QY 1801 LTTQNTLLFNILGQVAAQLAPPAAAFVAGIAGAAVSGIGLQKVLVDILAGYGAVA 1860
Db 1801 LTTQNTLLFNILGQVAAQLAPPAAAFVAGIAGAAVSGIGLQKVLVDILAGYGAVA 1860
QY 1861 GALVAFKMSGEVSTEDVNLNLLPAILSPGALVGVVCAAILRRHVGPGEAVQMMNRLI 1920
Db 1861 GALVAFKMSGEVSTEDVNLNLLPAILSPGALVGVVCAAILRRHVGPGEAVQMMNRLI 1920
QY 1921 AFASRGNHVSPTHVPSDAAARVTOILSSITITOLLKRLHQLWINECDSTPCSGSLRDV 1980
Db 1921 AFASRGNHVSPTHVPSDAAARVTOILSSITITOLLKRLHQLWINECDSTPCSGSLRDV 1980
QY 1981 WDMICTVLTDFTKTLQSKLLPRLPGVPLFCQRYGKGVWRGDIQMOTTCPCGAGIAGHVK 2040
Db 1981 WDMICTVLTDFTKTLQSKLLPRLPGVPLFCQRYGKGVWRGDIQMOTTCPCGAGIAGHVK 2040
QY 2041 NGSMRIIVGPRTCNSNTHGTFPINAYTTGCTPSPAPNYSRALWRVAEYEVTVRGDFH 2100
Db 2041 NGSMRIIVGPRTCNSNTHGTFPINAYTTGCTPSPAPNYSRALWRVAEYEVTVRGDFH 2100
QY 2101 YVTGMTDNNVCKQCPQVPAPEFTVDGVLRLHRYAPACKPLLRDVTFOVGLNQVLVGSQ 2160
Db 2101 YVTGMTDNNVCKQCPQVPAPEFTVDGVLRLHRYAPACKPLLRDVTFOVGLNQVLVGSQ 2160
QY 2161 PCEPEPDVVLTSMLTDSHITAEAKERLARGSPPSLASSASQLSAPSLKATCTTHHD 2220
Db 2161 PCEPEPDVVLTSMLTDSHITAEAKERLARGSPPSLASSASQLSAPSLKATCTTHHD 2220

QY 2221 SPADLIEANLLWRQMGNGNITRVESENKVVILDSPEPLHAEGDEREISVAAEILKRSK 2280
Db 2221 SPADLIEANLLWRQMGNGNITRVESENKVVILDSPEPLHAEGDEREISVAAEILKRSK 2280
QY 2281 FPSALPIWARPDPYNPLLESXDPDYVPVVGCPPLPYKAPPIPPPRKRTVWLITESV 2340
Db 2281 FPSALPIWARPDPYNPLLESXDPDYVPVVGCPPLPYKAPPIPPPRKRTVWLITESV 2340
QY 2341 SSALAEALATKTFGSSGSAVDSGTATALDPLASDDGDKGSDVESYSPMPLEGEPCDPL 2400
Db 2341 SSALAEALATKTFGSSGSAVDSGTATAPDQVSDDGKSDVESYSPMPLEGEPCDPL 2400
QY 2401 SDGSWSTVSEASESDVVCSSMSYTWTCALITPCAEESEKLPINPLNSLLRHNWYAT 2460
Db 2401 SDGSWSTVSEASESDVVCSSMSYTWTCALITPCAEESEKLPINPLNSLLRHNWYAT 2460
QY 2461 SRASLRQKVTFDRLQVLDHRYDVLKEMKAKASTVAKALLSIEACKLTPHSAKSKP 2520
Db 2461 SRASLRQKVTFDRLQVLDHRYDVLKEMKAKASTVAKALLSIEACKLTPHSAKSKP 2520
QY 2521 GYKAKVRNLSSRAVNHRSVWEDILEDTEPTDTTIMAKSEVFCVQPKGGRKPARLIV 2580
Db 2521 GYKAKVRNLSSRAVNHRSVWEDILEDTEPTDTTIMAKSEVFCVQPKGGRKPARLIV 2580
QY 2581 FPDGLVRCRMALYDVVSTLPQAVMGSSYGYQYSPQKQVRFVLNWTWKKKCPMGFSYDT 2640
Db 2581 FPDGLVRCRMALYDVVSTLPQAVMGSSYGYQYSPQKQVRFVLNWTWKKKCPMGFSYDT 2640
QY 2641 RCFDSTVTSIEDIRVESIYOCCLAPAEARQAIRSLTERLYIGGPLTNSGQNGCYRRCRA 2700
Db 2641 RCFDSTVTSIEDIRVESIYOCCLAPAEARQAIRSLTERLYIGGPLTNSGQNGCYRRCRA 2700
QY 2701 SGVLTTSCTGNTLCYLAATAACRAAKLQDCTMIVNGDLDLWVIESAGTQEDAAALAEPT 2760
Db 2701 SGVLTTSCTGNTLCYLAATAACRAAKLQDCTMIVNGDLDLWVIESAGTQEDAAALAEPT 2760
QY 2761 AMTRYAPPDPPQPEYDLELITSCSSNVSEVADHAGKRVYVLTTRDPTPLAARAAETAR 2820
Db 2761 AMTRYAPPDPPQPEYDLELITSCSSNVSEVADHAGKRVYVLTTRDPTPLAARAAETAR 2820
QY 2821 HTPTNSWLGNIIMVAPTLWARMLTHTFSSILLAOEQLEKALDQIYGACYSIEPLDLQ 2880
Db 2821 HTPTNSWLGNIIMVAPTLWARMLTHTFSSILLAOEQLEKALDQIYGACYSIEPLDLQ 2880
QY 2881 IIBELHGLSFTLESYSPGEINRVASCLKXGLVPPLETWHRARSVRKLLSOGGEAATC 2940
Db 2881 IIBELHGLSFTLESYSPGEINRVASCLKXGLVPPLETWHRARSVRKLLSOGGEAATC 2940
QY 2941 GRVLFNVAETKLTLPAAASQIDLSGMFVAGYSGSDIYHLSRARPRHFFLCILLISV 3000
Db 2941 GRVLFNVAETKLTLPAAASQIDLSGMFVAGYSGSDIYHLSRARPRHFFLCILLISV 3000
QY 3001 GVGYYLLPNR 3010
Db 3001 GVGYYLLPNR 3010

RESULT 4

POLG HCVTW STANDARD; PRT; 3010 AA.
AC P29845;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21); (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)]
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]
OS Hepatitis C virus (isolate Taiwan) (HCV).

481 DORPYCWHAPRCGVYV2ASQVCGPVYCFTEPSVWVGTTRSGVPTYSWGNETDVMLN 540
 481 DORPYCWHAPRCGVYV2ASQVCGPVYCFTEPSVWVGTTRSGVPTYSWGNETDVMLN 540
 541 NTRPPQGNWFGCTWMASTGFTKTCGPPONIGGVGNRTLICPTDCPRKPEATYTKCGSG 600
 541 NTRPPQGNWFGCTWMASTGFTKTCGPPONIGGVGNRTLICPTDCPRKPEATYTKCGSG 600
 601 PWTTPRCLVDYFVRLWHYPCITLNPFSFKVUMYGVGVHRLNAACNWTGRBSCNLEORDRS 660
 601 PWTTPRCLVDYFVRLWHYPCITLNPFSFKVUMYGVGVHRLNAACNWTGRBSCNLEORDRS 660
 661 ELSPLLLSTTEWQILPCAFITLPAISLGLIHLHQNIVDVQYLYGVGSASFVSPFAIKWEYIL 720
 661 ELSPLLLSTTEWQILPCAFITLPAISLGLIHLHQNIVDVQYLYGVGSASFVSPFAIKWEYIL 720
 721 LFLLLIADARVCACLWMMLIIAQBAMALENLVYLNAAVAGAGHIIISFLVFFCAAWYIKG 780
 721 LFLLLIADARVCACLWMMLIIAQBAMALENLVYLNAAVAGAGHIIISFLVFFCAAWYIKG 780
 781 RIAPGAYAFYGVWVPLLLALLALPPRAYALDREMAASCGGAVLVGVHFLTLSPYKVFELT 840
 781 RIAPGAYAFYGVWVPLLLALLALPPRAYALDREMAASCGGAVLVGVHFLTLSPYKVFELT 840
 841 RLIIWMLQYFTRABEAHQVWVPLNVNRRGRDAIILLTCAVHPELIPDITKLLAILGLPLM 900
 841 RLIIWMLQYFTRABEAHQVWVPLNVNRRGRDAIILLTCAVHPELIPDITKLLAILGLPLM 900
 901 VLQAGITRVFYFVRAOGLIIRACMLVRKVAGHYVQVFMKIGALTGYVYVNHILTPLEADWA 960
 901 VLQAGITRVFYFVRAOGLIIRACMLVRKVAGHYVQVFMKIGALTGYVYVNHILTPLEADWA 960
 961 HAGLEDIAVAVPVPVFMETKVTWADTAACGDIILGHPVSAARRKEIFLGPADSLGEG 1020
 961 HTGURDIAVAVPVPVFMETKVTWADTAACGDIILGHPVSAARRKEIFLGPADSLGEG 1020
 1021 QGWELLAPITAYSQOTRGVLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCINGVCWT 1080
 1021 QGWELLAPITAYSQOTRGVLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCINGVCWT 1080
 1081 VYHAGSKTLGPKGPTOMYTNVDLILVQWQAPPGARSMTPCSCGSSDIYLVTRHADVI 1140
 1081 VYHAGSKTLGPKGPTOMYTNVDLILVQWQAPPGARSMTPCSCGSSDIYLVTRHADVI 1140
 1141 PVRRGDSRGLSPRISYLGSGSGGLLCPSGHVGVFRAAVCTSGVAKAVDFIVPES 1200
 1141 PVRRGDSRGLSPRISYLGSGSGGLLCPSGHVGVFRAAVCTSGVAKAVDFIVPES 1200
 1201 METTMRSPVTDNSTPFAVPTQVVAHLHAPTGSKTKVPAAYAAQYKVLVLPNSVAA 1260
 1201 METTMRSPVTDNSTPFAVPTQVVAHLHAPTGSKTKVPAAYAAQYKVLVLPNSVAA 1260
 1261 TLGFCAYMSKAGHDPMIRTCVTRITTTGGSITYSYGKFLADGCGSGAYDIILICDECHS 1320
 1261 TLGFCAYMSKAGHDPMIRTCVTRITTTGGSITYSYGKFLADGCGSGAYDIILICDECHS 1320
 1321 TDSITLIGIGVLDQAEATAGARLVYLAATPPGVSVPVPHNIEIRIGLSNNGEIPFYCKAI 1380
 1321 TDSITLIGIGVLDQAEATAGARLVYLAATPPGVSVPVPHNIEIRIGLSNNGEIPFYCKAI 1380
 1381 PIRAIKGRHUIFCHSKKKDELAALKTGLGLNAVAYYRGLDVSVIPPIGDVVVATDAL 1440
 1381 PIRAIKGRHUIFCHSKKKDELAALKTGLGLNAVAYYRGLDVSVIPPIGDVVVATDAL 1440
 1441 MTGFTGDFSDVIDNCVTQVDFSLDPTFTLETTVPQDAVSRSORGRGRSGIYR 1500
 1441 MTGFTGDFSDVIDNCVTQVDFSLDPTFTLETTVPQDAVSRSORGRGRSGIYR 1500
 1501 FVTPGERPSGMPDSSVLCYDAGCAWYELTFAETSVRILRAYLNTPLPVQCQDHFLEWES 1560
 1501 FVTPGERPSGMPDSSVLCYDAGCAWYELTFAETSVRILRAYLNTPLPVQCQDHFLEWES 1560
 1561 VFTGLTHIDAHLFSLQKQAGNFPVLVAYQATVCARAQAPPPSWDMQWKCLIRLKPFTLHG 1620

1561 VFTGLTHIDAHLFSLQKQAGNFPVLVAYQATVCARAQAPPPSWDMQWKCLIRLKPFTLHG 1620
 1621 PPPLLYRIGAVONEVILLTHPIKYIMACMSADLEVVTTWLVGVGLAALAAAYCIUTTGVS 1680
 1621 PPPLLYRIGAVONEVILLTHPIKYIMACMSADLEVVTTWLVGVGLAALAAAYCIUTTGVS 1680
 1681 VTVGRIILSGKPAVVPDRREVLYQEFDEMEECASQLPYIEQGMQLAEOFKQKALGLQIAT 1740
 1681 VTVGRIILSGKPAVVPDRREVLYQEFDEMEECASQLPYIEQGMQLAEOFKQKALGLQIAT 1740
 1741 KOAEAAAAPVVEKWEALFTFWAKHWNFISGTOYLAGLSTLPCNFAIASLMAFTASITSP 1800
 1741 KOAEAAAAPVVEKWEALFTFWAKHWNFISGTOYLAGLSTLPCNFAIASLMAFTASITSP 1800
 1801 LTTONTLFIILGGVAAQOLAPPSAASAFVAGIAGAAVGSIGLGKVLVDIILAGVAGVA 1860
 1801 LTTONTLFIILGGVAAQOLAPPSAASAFVAGIAGAAVGSIGLGKVLVDIILAGVAGVA 1860
 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVDPOEGAVQVWNRLLI 1920
 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVDPOEGAVQVWNRLLI 1920
 1921 AFASRGNHVSPTHYVPESDAAAARVTOILLSLITITOLLKELHOMINEDCSTPCSGWLRDV 1980
 1921 AFASRGNHVSPTHYVPESDAAAARVTOILLSLITITOLLKELHOMINEDCSTPCSGWLRDV 1980
 1981 WDMICVILDFKTLQSLKLLPRLPGVFFLSCORGKGVWVRGDIIMOTTCPCCAQIAGHVK 2040
 1981 WDMICVILDFKTLQSLKLLPRLPGVFFLSCORGKGVWVRGDIIMOTTCPCCAQIAGHVK 2040
 2041 NGSMRIVGSPRTCSNTWHGTFPINAYTTGCTPSPAPNYSRALRVAABEYVEVTRVGDH 2100
 2041 NGSMRIVGSPRTCSNTWHGTFPINAYTTGCTPSPAPNYSRALRVAABEYVEVTRVGDH 2100
 2101 YVTGMTTDNVKPCQVPAPEFFTEVDGVLHRYAPACKELLREDYTFQVGLQYLVGSOL 2160
 2101 YVTGMTTDNVKPCQVPAPEFFTEVDGVLHRYAPACKELLREDYTFQVGLQYLVGSOL 2160
 2161 PCPEPDVTVLTQMLTDPGSHITAEATKRRLARGSPPLSLASSASQLSAPSLKATCTTHD 2220
 2161 PCPEPDVTVLTQMLTDPGSHITAEATKRRLARGSPPLSLASSASQLSAPSLKATCTTHD 2220
 2221 SPADLIEANLLWRQMGNGNITRVESKENVILDSPEPHABEGDBREISVAAIRLKRSK 2280
 2221 SPADLIEANLLWRQMGNGNITRVESKENVILDSPEPHABEGDBREISVAAIRLKRSK 2280
 2281 FPSALPIWARPDVNPPLLESWKDPDVPVWVHGCPLPPTKAPPIPPPRKRTVVLTESNV 2340
 2281 FPSALPIWARPDVNPPLLESWKDPDVPVWVHGCPLPPTKAPPIPPPRKRTVVLTESNV 2340
 2341 SSALAEIATKTFGSSGSSAVDSGTATAPDLASDDGDKGSDVESYSSMPLEGEPPDL 2400
 2341 SSALAEIATKTFGSSGSSAVDSGTATAPDLASDDGDKGSDVESYSSMPLEGEPPDL 2400
 2401 SDGSWSTVSEBSEDVCCSNSTWTGALITPCAAESKLPINPIGNSLRLHNNVYATT 2460
 2401 SDGSWSTVSEBSEDVCCSNSTWTGALITPCAAESKLPINPIGNSLRLHNNVYATT 2460
 2461 SRASLROKVTDFRLQVLDHVRDLKEMKAKASTVKAKLISIEBACKLTPPHSARSKF 2520
 2461 SRASLROKVTDFRLQVLDHVRDLKEMKAKASTVKAKLISIEBACKLTPPHSARSKF 2520
 2521 GYGAKOYRNLSSRAVNIHRSWEDLLEDETETPIDTTIMAKSEVFCVQPKGGRKPARLIV 2580
 2521 GYGAKOYRNLSSRAVNIHRSWEDLLEDETETPIDTTIMAKSEVFCVQPKGGRKPARLIV 2580
 2581 PFDLGVVVCMAIYDVVSTLPOAVMGSSYGFQYSPKQVREFLVNTWKSXCCPMGYSYDT 2640
 2581 PFDLGVVVCMAIYDVVSTLPOAVMGSSYGFQYSPKQVREFLVNTWKSXCCPMGYSYDT 2640
 2641 RCFDSTVTESDIRVEESIOCCDILAPEARQAIRSLTERLYIGGPLTNSKQNCQYRRCRA 2700

Db 2641 RCFDSTVETSDIRVEESYQCCDLAPARQAIRSUTERLYIGGFLTNKSGQVQYRCRA 2700

Qy 2701 SGVLTTCGNTLTCYLKATACRAKLODCTMYNGVDDLVICISAGFOEDAAALRAFTE 2760

Db 2701 SGVLTTCGNTLTCYLKASACRAKLODCTMYNGVDDLVICISAGFOEDAAALRAFTE 2760

Qy 2761 ANTRYSAPPGPDPPEYDELEITSCSSNVSAHDASGKRVYLLTRDPTPLARAAMETAR 2820

Db 2761 ANTRYSAPPGPDPPEYDELEITSCSSNVSAHDASGKRVYLLTRDPTPLARAAMETAR 2820

Qy 2821 HTPNSWLGNTIMVAPTWARMLMTHPFSILLAOEQLKALDCQIYACYSIEPLDLPQ 2880

Db 2821 HTPNSWLGNTIMVAPTWARMLMTHPFSILLAOEQLKALDCQIYACYSIEPLDLPQ 2880

Qy 2881 IIERHLGSLAFTLSYSGBEINRVASCLRGVPLRTWRHARSVRKLLSQGGRATC 2940

Db 2881 IIERHLGSLAFTLSYSGBEINRVASCLRGVPLRTWRHARSVRKLLSQGGRATC 2940

Qy 2941 GRYLFNNAVTKLTPIPAASQDLSCWFWAGYSGGDIYHLSRARPWFPLCLLLLSV 3000

Db 2941 GRYLFNNAVTKLTPIPAASQDLSCWFWAGYSGGDIYHLSRARPWFPLCLLLLSV 3000

Qy 3001 GVGIVYLPNR 3010

Db 3001 GVGIVYLPNR 3010

RESULT 5

ID POLG_HCV1 STANDARD; PRT; 3011 AA.

AC P26664;

DT 01-AUG-1992 (Rel. 23, Created)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P66); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).

OS Hepatitis C virus (isolate 1) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OX NCBI_TaxID=11104;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=91172826; PubMed=1848704;

RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;

RT "Genetic organization and diversity of the hepatitis C virus."; Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

RL -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA}(N).

CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC EMBL; M62321; AAA45676.1; --

DR PIR; A39166; GNWVC3.

DR PDB; 1AIY; 16-FEB-98.

DR PDB; 1HEI; 25-NOV-98.

DR MEROPS; S29.001; --

DR MEROPS; U39.001; --

DR InterPro; IPR009003; Cys_Ser_trypsin.

DR InterPro; IPR001410; DBAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV_env.

DR InterPro; IPR002531; HCV_NS1.

DR InterPro; IPR002518; HCV_NS2.

DR InterPro; IPR000745; HCV_NS4a.

DR InterPro; IPR001490; HCV_NS4b.

DR InterPro; IPR002868; HCV_NS5a.

DR InterPro; IPR002166; HCV_RdRp.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR004109; Peptidase_C29.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01543; HCV_capsid_1.

DR Pfam; PF01542; HCV_core_1.

DR Pfam; PF01539; HCV_env_1.

DR Pfam; PF01560; HCV_NS1_1.

DR Pfam; PF01538; HCV_NS2_1.

DR Pfam; PF02907; HCV_NS3_1.

DR Pfam; PF01006; HCV_NS4a_1.

DR Pfam; PF01001; HCV_NS4b_1.

DR Pfam; PF01506; HCV_NS5a_1.

DR Pfam; PF00271; Helicase_C_1.

DR Pfam; PF00998; Viral_RGRP_1.

DR ProDom; PD186062; HCV_NS1_1.

DR SMART; SM00487; DEXDc_1.

DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;

KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;

KW 3D-structure.

FT INIT_MFT 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.

FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).

FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).

FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT ACT_SITE 1083 1083 POTENTIAL.

FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP_BIND 1230 1237 ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 136 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCEB5AF9 CRC64;
Query Match 88.2%; Score 14119.5; DB 1; Length 3011;
Best Local Similarity 85.2%; Pred. No. 0;
Matches 2564; Conservative 226; Mismatches 220; Indels 1; Gaps 1;
QY 1 MSTNPKCRKXKNTNRPPQDVKPGGGQIVGGYVLLPRGPRGLGVRAATKASRSQPRG 60
DB 1 MSTNPK3QKKNKNTNRPPQDVKPGGGQIVGGYVLLPRGPRGLGVRAATKTSRSQPRG 60
QY 61 RROQIPKARPEGRAWAQPGYFWPLYNGLNAGWLLSPRGSRPSWGPDPERRSRNLG 120
DB 61 RROQIPKARPEGRAWAQPGYFWPLYNGLNAGWLLSPRGSRPSWGPDPERRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVVLEDCVNTATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVVLEDCVNTATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASAYERNVSGIHYVNDSCNSSIIVRAADVIMHTPCVPCVQBGNSRCWV 240
DB 181 LLSCLTTPASAYERNVSGIHYVNDSCNSSIIVRAADVIMHTPCVPCVQBGNSRCWV 240
QY 241 ALPTLAARNASVPTTIRRHWDLLVGTAAFCSAMVYVGDLCGSIPLYSQLFPTSPRRHET 300
DB 241 AMTPTVATROGKLPATQIRHIDLLVGSATLCSALYVGDLCGSVFLVQOLFPTSPRRHET 300
QY 301 VQDNCSTYPGHVGSHRMANDMNWNWPTTALVVSQILLRIPQAVVDMVAGHWGLAGLA 360
DB 301 TOGNCSTYPGHVGSHRMANDMNWNWPTTALVMAQLIRIQAILLMIAGHWGLAGIA 360
QY 361 YYSVVGWNAKVLIVALLFAGVDGTHTRVAGHTTSGTSLFSSGASQKQLVNTNGSW 420
DB 361 YFSVVGWNAKVLIVALLFAGVDGTHTRVAGHTTSGTSLFSSGASQKQLVNTNGSW 420
QY 421 HINRTALNCNDSLOTGFAALFYAHKFNSCCPERMASCRIDWFAQGWGPITVTKPNSS 480
DB 421 HLNSTALNCNDSLNTGWLAGLYFHHKFNSCCPERLASCRPLTDFDQNGGPISVANGSGP 480
QY 481 DQRYCWHYAPRCGVVPASQVCGPVYCFPTSPVVGTTDSSGVPYTSWGNEDTVMMLN 540
DB 481 DQRYCWHYAPRCGVVPASQVCGPVYCFPTSPVVGTTDSSGVPYTSWGNEDTVMMLN 540
QY 541 NTRPPQGNWFGCTWNSGTGKTGCGPPCNIGGVGNRLTLCPTDCFRKHPEATVTKCGSG 600
DB 541 NTRPPQGNWFGCTWNSGTGKTGCGPPCNIGGVGNRLTLCPTDCFRKHPEATVTKCGSG 600
QY 601 PWTLPRLVDYPRILWHPYCTLNFSIFKVRMYGVGVZHRLEAANWTRGERCNLEDRDRS 660
DB 601 PWTLPRLVDYPRILWHPYCTLNFSIFKVRMYGVGVZHRLEAANWTRGERCNLEDRDRS 660
QY 661 ELSPLLLSTTWOILPCAPFTLIPALSTGLIHLHONIYDVQVLYGVGSAFVSPALKWEYIL 720
DB 661 ELSPLLLSTTWOILPCAPFTLIPALSTGLIHLHONIYDVQVLYGVGSAFVSPALKWEYIL 720
QY 721 LFLLLADARVACACMLMMLIAQAEALLENVLNAAVSAGAHGTLSPLVFFCAAWIKG 780
DB 721 LFLLLADARVACACMLMMLIAQAEALLENVLNAAVSAGAHGTLSPLVFFCAAWIKG 780
QY 781 RLAFCAAYAFYGVWPELLELLALPPRAYALDREMAASCGGAVLGLVFLTSLSPYKVFELT 840
DB 781 KWPVCAVTFYGVWPELLELLALPPRAYALDREMAASCGGAVLGLVFLTSLSPYKVFELT 840
QY 841 PLIWLQVFIIRAHMOVWVPLNVRGGRDAIILLTCAVHPBELFDITKILLAILGLPILM 900
DB 841 WCLMWLQVFLTRVQAQLEWVPLNVRGGRDAIILLTCAVHPBELFDITKILLAILGLPILM 900
QY 901 VLOAGITRVPVVRQAQLEWVPLNVRGGRDAIILLTCAVHPBELFDITKILLAILGLPILM 960
DB 901 ILQASLLKVPVVRQAQLEWVPLNVRGGRDAIILLTCAVHPBELFDITKILLAILGLPILM 960

QY 961 HAGLRDLAVAVPVPFVSAMETKVIITWCADTAACGDIILGLPVSARRKBIPLGPADSLG 1020
DB 961 HNGLRDLAVAVPVPFVSAMETKVIITWCADTAACGDIILGLPVSARRKBIPLGPADSLG 1020
QY 1021 QGWRLLAPITAYSQOTRQGVGLGCIITSLTRDRKNQVEGEVQVSTATQSFACINVCVWT 1080
DB 1021 KGWRLAPITAYSQOTRQGVGLGCIITSLTRDRKNQVEGEVQVSTATQSFACINVCVWT 1080
QY 1081 VYHGASKTLAPKGPBITOMYTNVDLIDVWQAPPCGARSMTFCSCGSSDLYLVTRHADVI 1140
DB 1081 VYHGASKTLAPKGPBITOMYTNVDLIDVWQAPPCGARSMTFCSCGSSDLYLVTRHADVI 1140
QY 1141 PVRRRSGDSSLLSPFVSYLKSGSGCLLPCPSGHVGVVFRAAVCTRGVAKAVDFIPVES 1200
DB 1141 PVRRRSGDSSLLSPFVSYLKSGSGCLLPCPSGHVGVVFRAAVCTRGVAKAVDFIPVES 1200
QY 1201 METTMSPVFTDNSTPPAPVPTQVQVHLHAPTPGSGKSTKVPAAYAAQKVLVLPNSVAA 1260
DB 1201 LETTMSPVFTDNSTPPAPVPTQVQVHLHAPTPGSGKSTKVPAAYAAQKVLVLPNSVAA 1260
QY 1261 TLGFGAYMSKAGIGDNIRIGVRTITTTGSGITVYTKPLADGGCGSGGAYDIIICDECHS 1320
DB 1261 TLGFGAYMSKAGIGDNIRIGVRTITTTGSGITVYTKPLADGGCGSGGAYDIIICDECHS 1320
QY 1321 TDSITLIGITVLDQAETAGARLVWLATATPPGQSVTPHPNIBEEICLSNNGEIPFYKAI 1380
DB 1321 TDSITLIGITVLDQAETAGARLVWLATATPPGQSVTPHPNIBEEICLSNNGEIPFYKAI 1380
QY 1381 PIALKGRHLIFCHSKKCCDELAALGILGNVAVYAGLDVSVIPIPGDVVVVATDAL 1440
DB 1381 PLEVIKGRHLIFCHSKKCCDELAALGILGNVAVYAGLDVSVIPIPGDVVVVATDAL 1440
QY 1441 MTGFTGDFDVIDCNTCVTQVDFSLDPTTETTTVPQDAVSRSORRGRTGRSGIYR 1500
DB 1441 MTGFTGDFDVIDCNTCVTQVDFSLDPTTETTTVPQDAVSRSORRGRTGRSGIYR 1500
QY 1501 FVTPGERPSGMPDSSVLCYCYDAGCAWELTPAETSVRLRAYLNTPLPCVQDHLFEWES 1560
DB 1501 FVTPGERPSGMPDSSVLCYCYDAGCAWELTPAETSVRLRAYLNTPLPCVQDHLFEWES 1560
QY 1561 VFTGLTHIDAHFLSOTKQAGDNFYLVAQATVCARAQAPPPSDQWKKLILKPTLHG 1620
DB 1561 VFTGLTHIDAHFLSOTKQAGDNFYLVAQATVCARAQAPPPSDQWKKLILKPTLHG 1620
QY 1621 PTLKXKLGAVQNEVILTHPTKYIMACMSADLEWVTSWVLVGGVLAALAAVCLATGCV 1680
DB 1621 PTLKXKLGAVQNEVILTHPTKYIMACMSADLEWVTSWVLVGGVLAALAAVCLATGCV 1680
QY 1681 VIVGRVILSGKPAIIPREVLYREFDEMEBCSQHLPYIEQGMMAEQFKQKALGLLQAT 1740
DB 1681 VIVGRVILSGKPAIIPREVLYREFDEMEBCSQHLPYIEQGMMAEQFKQKALGLLQAT 1740
QY 1741 KQAEAAAPVVBESKWRALETFMAKHMNFISGIOYLAGISTLPCNPATIASIAETASTSP 1800
DB 1741 KQAEAAAPVVBESKWRALETFMAKHMNFISGIOYLAGISTLPCNPATIASIAETASTSP 1800
QY 1801 LITQNTLLFNILGOWAAQALAPPASAFVAGIAGAAVGSIGLGKVLVDILAGYGAGVA 1860
DB 1801 LITQNTLLFNILGOWAAQALAPPASAFVAGIAGAAVGSIGLGKVLVDILAGYGAGVA 1860
QY 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEVQVWNRLLI 1920
DB 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEVQVWNRLLI 1920
QY 1921 APASRGNHVSPTHVVPESDAAAVTOLLSLTITQLLRHLOWINBEDCSTPCSGSWLRDV 1980
DB 1921 APASRGNHVSPTHVVPESDAAAVTOLLSLTITQLLRHLOWINBEDCSTPCSGSWLRDV 1980
QY 1981 WDMITCVLTDFKTLQSKLPLRPGVFFLSCQRCQYKGVWFGDGMOTTCPCGAQIAGHVK 2040
DB 1981 WDMITCVLTDFKTLQSKLPLRPGVFFLSCQRCQYKGVWFGDGMOTTCPCGAQIAGHVK 2040

DR TRANSFAC; T04155; --
 DR InterPro; IPR009003; Cys Ser_trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR004109; Peptidase C29.
 DR InterPro; IPR007095; RNA pol DS ps.
 DR InterPro; IPR007094; RNA pol_PsVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 192 383 CAPSID PROTEIN C.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 810 1026 PROTEIN P7.
 FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
 FT CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5B.
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1224 1226
 FT TURN 1232 1233
 FT TURN 1236 1238
 FT TURN 1239 1246
 FT TURN 1247 1248
 FT STRAND 1251 1255
 FT HELIX 1258 1271

1272 1272
 1277 1280
 1281 1282
 1283 1285
 1291 1295
 1296 1301
 1302 1303
 1312 1316
 1317 1319
 1323 1335
 1336 1340
 1343 1347
 1352 1353
 1360 1361
 1362 1366
 1368 1368
 1373 1375
 1376 1377
 1378 1380
 1382 1385
 1389 1393
 1397 1403
 1410 1411
 1414 1417
 1419 1420
 1432 1436
 1438 1439
 1450 1453
 1456 1463
 1471 1478
 1480 1480
 1481 1488
 1489 1490
 1497 1501
 1507 1507
 1511 1511
 1514 1527
 1532 1544
 1550 1550
 1555 1564
 1570 1578
 1579 1580
 1584 1597
 1598 1598
 1606 1611
 1614 1618
 1622 1623
 1627 1627
 1635 1636
 1640 1652
 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
 SQ SEQUENCE
 Query Match 87.0%; Score 13935.5; DB 1; Length 3011;
 Best Local Similarity 84.5%; Pred.No. 0;
 Matches 254; Conservative 226; Mismatches 239; Indels 1; Gaps 1;
 Qy 1 MSTNPKPQRXTKNTNRRPQDVKFPGGQIVGGVYLLPRRGPRLGVRAIKASERQPRG 60
 1 MSINPKPQRXTKNTNRRPQDVKFPGGQIVGGVYLLPRRGPRLGVRAIKASERQPRG 60
 Qy 61 RRQIPKAREPEGRANAQPGYPWPLYGNEGLGWAGWMLSPRGRPSWGPTDPRRRNLG 120
 61 RRQIPKAREPEGRNTWAQPGYPWPLYGNECGWAGWMLSPRGRPSWGPTDPRRRNLG 120
 Qy 121 KVIDTLTCGFADLMGYPLVGLGALGGAALAHGVLELDGVNYATGNLPGCCFSIFLLA 180
 121 KVIDTLTCGFADLMGYPLVGLGALGGAALAHGVLELDGVNYATGNLPGCCFSIFLLA 180
 Qy 181 LLSCLTIPASAYEVNVSGLYHYVTNDCSNSSIVYEADVIMHTFGVPCVQEGNSRCWV 240
 181 LLSCLTVPASAYQVRNSSLGLYHYVTNDCPNSSVVVEADALHTPGVPCVREGNASRCWV 240
 Qy 241 ALTFTLAARNASVFTTIRRHVDLLVGTAAFCGAMVVDLCGSIPLVSQLTFTSPRHET 300

Db 241 AVIPTVATRDGKLPPTQERHDLVGSATLCSALYVGLCGSVLFGQLFTFSPRHWT 300
Qy 301 VODCNCSYPGHVSGRHAWMMNWSPTTALVVSOLLARIPOAVVDMVAGHGWGLAGLA 360
Db 301 TODCNCSYPGHITGRHAWMMNWSPTTALVVSOLLARIPOAIMDIAGHGWGLAGK 360
Qy 361 YSMYGNNAKVLIVALLFAGVDGETHHTEKRVAGHTTSFSTLSSGASOKIQLVNTNSW 420
Db 361 YPSMYGNNAKVLIVALLFAGVDAETHVTCGNAGRTTAGLVGLTFPGAKQNIQLVNTNSW 420
Qy 421 HINRTALNCNDSLOTGPPAALFYAHKFNSGCEBMSRCPDWAQOGPITYTKPNSS 480
Db 421 HINRTALNCNDSINTGWLAGFYQHKFNSGCEBMSRCPDWAQOGPITYTKPNSS 480
Qy 481 DORPYCWHYAPFCGVVPASQVCGPVYCTPFPVVGTTDRSGVPTYSNGENETDVLIN 540
Db 481 DRPYCWHYPPPCGVIPAKSVCGPVYCTPFPVVGTTDRSGAPYISGANDTDVFLIN 540
Qy 541 NTRPPQGNWFGTWMNSTGFTKOCGPPCNIGVGNRTLICPTDCFRKHPEATYKCGSG 600
Db 541 NTRPPQGNWFGTWMNSTGFTKOCGAPPVIGVGNRTLICPTDCFRKHPEATYKCGSG 600
Qy 601 PWLTERCLVDYRLWHYPCNTLNFSTFKYNTVGGVHEHLNAACNWTGRCCNLEDRDS 660
Db 601 PRITPRCMVDYRPLWHYPCNTLNFSTFKYNTVGGVHEHLNAACNWTGRCCNLEDRDS 660
Qy 661 ELSPLLSSTEQIILPCAFTTLPALSTGLIHLHQNIVDVQYLVGYSAPVSPAIKWEYIL 720
Db 661 ELSPLLSSTEQIILPCAFTTLPALSTGLIHLHQNIVDVQYLVGYSAPVSPAIKWEYIL 720
Qy 721 LFLILLADARVACACIWMMLIIQAARALNVLVNAASVAGAHGILSFLVFFCAAWYIKG 780
Db 721 LFLILLADARVACACIWMMLIIQAARALNVLVNAASVAGAHGILSFLVFFCAAWYIKG 780
Qy 781 RLAPCAAFYGVWVPLILLALPRAVALDEMAASCGGAVLGLVLTLSPPYKVELT 840
Db 781 RWPDAVVALYGNWVPLILLALPRAVALDEMAASCGGAVLGLVLTLSPPYKVELT 840
Qy 841 RLIMWLOYFITRAEAHMOWVPLNVGRGORDAILITCAVHPELLFDITKILLAILGLPM 900
Db 841 WCMWMLQYFLTEVEAQLHWVPLNVGRGORDAILITCAVHPELLFDITKILLAILGLPM 900
Qy 901 VLOAGITVPVPAQGLIRACMLVRKVAGHYVQVFMKLGALGTVYVNHLEPLRDA 960
Db 901 ILOAGLLVPVVRVQGLIRACMLVRKVAGHYVQVFMKLGALGTVYVNHLEPLRDA 960
Qy 961 HAGRLDVAVEPVVFSAMETKVIWAGADTAACGDIILGLPVSARRKEIFLGRADSLEG 1020
Db 961 HNGRLDVAVEPVVFSAMETKVIWAGADTAACGDIILGLPVSARRKEIFLGRADSLEG 1020
Qy 1021 QGWRLLAPITAYSQOTRGVLGCIITSLTRDRKNQVEGEVQVSTATQSLATCINGVCWT 1080
Db 1021 KGRLLAPITAYSQOTRGVLGCIITSLTRDRKNQVEGEVQVSTATQSLATCINGVCWT 1080
Qy 1081 VYHAGSXTLAPKPKPIQMTYNVLDLVGWAQPECARSMTPCSCGSDLLVYTHADVI 1140
Db 1081 VYHAGSXTLAPKPKPIQMTYNVLDLVGWAQPECARSMTPCSCGSDLLVYTHADVI 1140
Qy 1141 PVRARGDSRGLLSRPPVSYLKGSSGGPLLCPSGHVGVVERAAVCTRGVAKAVDTPVES 1200
Db 1141 PVRARGDSRGLLSRPPVSYLKGSSGGPLLCPSGHVGVVERAAVCTRGVAKAVDTPVES 1200
Qy 1201 METNRSVPFTDNTSPVAVPQTFQVAHLHAPTGSOKSTKVPAAYAAQYKVLVLPNSVAA 1260
Db 1201 LETNRSVPFTDNTSPVAVPQTFQVAHLHAPTGSOKSTKVPAAYAAQYKVLVLPNSVAA 1260
Qy 1261 TLGFGAYMSKAGIDPNRTGTGRTTTCGSITYSYGKFLADGCGSGAYDIIICDECHS 1320
Db 1261 TLGFGAYMSKAGIDPNRTGTGRTTTCGSITYSYGKFLADGCGSGAYDIIICDECHS 1320
Qy 1321 TDSITLIGTIGTLDQAEATAGARLVVLAATATPPGVSVTVPHPNEIEIGLSNNGEIPYKAI 1380
Db 1321 TDSITLIGTIGTLDQAEATAGARLVVLAATATPPGVSVTVPHPNEIEIGLSNNGEIPYKAI 1380
Qy 1381 PLEAKGRGHLIFCHSKKKKDBLAUKTLGLINAVAYRGLDVSVIPIGDVVVVATDAL 1440
Db 1381 PLEVAKGRGHLIFCHSKKKKDBLAUKTLGLINAVAYRGLDVSVIPITSGDVVVVSTDAL 1440
Qy 1441 MTGFTGDPDSVIDCNCVTQTVDPSLDPTFTIETTTVPQDAVSRSQRGRGTGRSGIYR 1500
Db 1441 MTGFTGDPDSVIDCNCVTQTVDPSLDPTFTIETTTVPQDAVSRSQRGRGTGRSGIYR 1500
Qy 1501 FVTPGERSGMFDSVLCCEYDAGCAWYELTTPAETSUMLRAYLNTPGULPVQDHLFVWS 1560
Db 1501 FVTPGERSGMFDSVLCCEYDAGCAWYELTTPAETSUMLRAYLNTPGULPVQDHLFVWS 1560
Qy 1561 VFTGLTHDAHFLSOTKQAGDNFPYLVAYQATVCARACAPPSWDMKCLIRLKTPLUG 1620
Db 1561 VFTGLTHDAHFLSOTKQAGDNFPYLVAYQATVCARACAPPSWDMKCLIRLKTPLUG 1620
Qy 1621 PTELLYRLGAVQNEVILTHPIITKYIMACSADLEVTSTWLVGVVLAALAAAYCITGTSV 1680
Db 1621 PTELLYRLGAVQNEVILTHPIITKYIMACSADLEVTSTWLVGVVLAALAAAYCITGTSV 1680
Qy 1681 VIVGRILISGKPAVVPDREVLYQFDEMEBCASQULPIYEQGKOLAEOPKQKALGLLOTAT 1740
Db 1681 VIVGRILISGKPAVVPDREVLYQFDEMEBCASQULPIYEQGKOLAEOPKQKALGLLOTAT 1740
Qy 1741 KOAABAAAPWESKRALETFWAKHWNFTSGIOYLAGLSTLPGNPAIASLMAFTASITSP 1800
Db 1741 RHAETVITAVQNTWQKLEVFVWAKHWNFTSGIOYLAGLSTLPGNPAIASLMAFTASITSP 1800
Qy 1801 LTTQNTLILFNILGGWAAQLAAPGAATAFVAGLAGLSTLPGNPAIASLMAFTASITSP 1860
Db 1801 LTTQNTLILFNILGGWAAQLAAPGAATAFVAGLAGLSTLPGNPAIASLMAFTASITSP 1860
Qy 1861 GALVAFKMSGEVSTEDLVNLLPAILSPGALVGVVCAATILRRHVGEVGEVQVMMNELI 1920
Db 1861 GALVAFKMSGEVSTEDLVNLLPAILSPGALVGVVCAATILRRHVGEVGEVQVMMNELI 1920
Qy 1921 AFASRGHVSPTHVPESDAAARVTQILSSLTITQKRLHQTWINECOSTPCSSWLRDV 1980
Db 1921 AFASRGHVSPTHVPESDAAARVTQILSSLTITQKRLHQTWINECOSTPCSSWLRDV 1980
Qy 1981 WDWITVLTDPKTLWQKLLPRLPGVPPPLSCQRYGKVGWGDGIMOTTCPCGAOIAHVK 2040
Db 1981 WDWITVLTDPKTLWQKLLPRLPGVPPPLSCQRYGKVGWGDGIMOTTCPCGAOIAHVK 2040
Qy 2041 NGSMRIYGPRTCSNTHGTPINAVTTGCTPSPADNYSRALWVAAEYVUVTRVGDH 2100
Db 2041 NGSMRIYGPRTCSNTHGTPINAVTTGCTPSPADNYSRALWVAAEYVUVTRVGDH 2100
Qy 2101 YVTGTTDNDVRCPCQVPAPEFTTEVDGVRHRYAPACKPLREDVTTFQVGLNQYLVSQL 2160
Db 2101 YVTGTTDNDVRCPCQVPAPEFTTEVDGVRHRYAPACKPLREDVTTFQVGLNQYLVSQL 2160
Qy 2161 PCEBPPDVLTSLMTDPSHITATKRRRLARGSPPLSSASSASQLSAPSKATCTHHD 2220
Db 2161 PCEBPPDVLTSLMTDPSHITATKRRRLARGSPPLSSASSASQLSAPSKATCTHHD 2220
Qy 2221 SPDAHLTEANLLWRQENGNITRVESKENVILDSFPLHABGDEFRTISVAABTLKRSK 2280
Db 2221 SPDAHLTEANLLWRQENGNITRVESKENVILDSFPLHABGDEFRTISVAABTLKRSK 2280
Qy 2281 FPSALPWARPDYNPPLLESKMDPYVPPVHVHGCLEPTKAPPTPPRRKRTVVLTSNV 2340
Db 2281 PAPALPVAREDYNPLLVETWKKDYBPPVHVHGCLEPTKAPPTPPRRKRTVVLTSNV 2340
Qy 2341 SSALAEATKTFGSSGSAVDSGTATALPDIASDDGKGDGSDVESYSSMPLEGPDPDL 2400
Db 2341 PTALAEATKTFGSSGSAVDSGTATALPDIASDDGKGDGSDVESYSSMPLEGPDPDL 2400
Qy 2401 SDGWSVTSVEEA-SEDEVVCCSMSYMTGALITPCAAEESKLPTNPLNSILRHHNMVYAT 2460
Db 2401 SDGWSVTSVEEA-SEDEVVCCSMSYMTGALITPCAAEESKLPTNPLNSILRHHNMVYAT 2460

QY 2460 TSRSASLRQKVTDRDLQVLDDHYRDLVKEMKAKASTWKAALLSIEBACKLTPHPSAKSK 2519
 DB 2461 TSRSACQKVKVTDRQLQVLDLSDYQDVLKVEYKAASVKANLLSVEEACSLAPHSASKS 2520
 QY 2520 FGYGAKDVNLSRAVNHRSVWDLLEDTPTDTTITMAKSEVFCVQPEKGGKRPRLI 2579
 DB 2521 FGYGAKDVCHARKAVAHINSVWDLLEDSVTPDTTITMAKNEVFCVQPEKGGKRPRLI 2580
 QY 2580 VFPDLGVRCVKALYDVWSTLPOAVMGSSYGFQSPKQRPVETLVNTWTKSKCQWGPSYD 2639
 DB 2581 VFPDLGVRCVKALYDVWSTLPOAVMGSSYGFQSPKQRPVETLVNTWTKSKCQWGPSYD 2640
 QY 2640 TRCFDSTVTSRDIRVESIYQCCDLAPPEARQAISLTERLVIIGGLTNSKGQNGCYRRCR 2699
 DB 2641 TRCFDSTVTSRDIRVESIYQCCDLAPPEARQAISLTERLVIIGGLTNSKGQNGCYRRCR 2700
 QY 2700 ASGVLTTSQGNLTTCYLKATRAACRAKIQDCTMVLNGDGLVVICESAGTQSDAAALRAFT 2759
 DB 2701 ASRVLTTSQGNLTTCYLKATRAACRAKIQDCTMVLNGDGLVVICESAGTQSDAAALRAFT 2760
 QY 2760 EAMTRYAPPDPPQPEYDLELITSCSNVSVAHDAGKRVYVLTDRPTTTLARAAMETA 2819
 DB 2761 EAMTRYAPPDPPQPEYDLELITSCSNVSVAHDAGKRVYVLTDRPTTTLARAAMETA 2820
 QY 2820 RHTPINSWLGNIIMYAPTLWARMILMTHFFSILLAOEKLXDCQIYGACYSIEPLDLP 2879
 DB 2821 RHTPINSWLGNIIMYAPTLWARMILMTHFFSILLAOEKLXDCQIYGACYSIEPLDLP 2880
 QY 2880 QIIRLHGLSAFTLHSPGINSVACLRKLGYPPLTWHRARSVRKLLSOGGGAAT 2939
 DB 2881 PIIOQLHGLSAFTLHSPGINSVACLRKLGYPPLTWHRARSVRKLLSOGGGAAT 2940
 QY 2940 CGRYLFNMAVTKLPIPAASQDLSCWFWAGYSGGDIYHLSRAPRPFPLCLLLS 2999
 DB 2941 CGRYLFNMAVTKLPIPAASQDLSCWFWAGYSGGDIYHLSRAPRPFPLCLLLS 3000
 QY 3000 VGVGIYLLPNR 3010
 DB 3001 AGVGIYLLPNR 3011

RESULT 7
 POLG HCVJ6
 AC P26660; STANDARD; PRT: 3033 AA.
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus [isolate HC-J6] (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.,
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions."
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA}(N).
 -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 lipoprotein envelope. The envelope consists of two proteins:
 protein M and glycoprotein E. The nucleocapsid is a complex of
 protein C and mRNA.
 -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D00944; BAA00792.1; -.
 DR PIR; JQ1303; JQ1303.
 DR HSP; P27958; IHEI.
 DR MROPS; S29.001; -.
 DR MROPS; U39.001; -.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT SITE 1087 1087
 FT ACT SITE 1111 1111
 FT ACT SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 733
 CHAIN 734 1010
 CHAIN 1011 1619
 CHAIN 1620 1866
 CHAIN 1867 2017
 CHAIN 2018 3033
 TRANSMEM 347 369
 ACT SITE 1087 1087
 ACT SITE 1111 1111
 ACT SITE 1169 1169
 NP_BIND 1234 1241
 SITE 1320 1323
 CARBOHYD 196 196
 CARBOHYD 209 209
 CAPSID PROTEIN C (POTENTIAL).
 MATRIX PROTEIN (POTENTIAL).
 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 PROTEASE/HELICASE NS3 (POTENTIAL).
 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 POTENTIAL.
 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CHARGE RELAY SYSTEM (BY SIMILARITY).
 ATP (POTENTIAL).
 DECH BOX.
 N-LINKED (GLCNAC...) (POTENTIAL).
 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 75.2%; Score 12036.5; DB 1; Length 3033;
 Best Local Similarity 71.9%; Pred. No. 0;
 Matches 2189; Conservative 338; Mismatches 472; Indels 45; Gaps 7;

Qy 1 MSTNPKQKRTNRRPQVPPGGQIYGVYLLPRRGRGLGVRAIRKASERSQPRG 60
 Db 1 MSTNPKQKRTNRRPQVPPGGQIYGVYLLPRRGRGLGVRAIRKASERSQPRG 60

Qy 61 RROPTKARREPGRAWQPGYPMPLYGNEGGLGAGWLLSPRSPSGPTDPRRSNIG 120
 Db 61 RROPTKARREPGRAWQPGYPMPLYGNEGGLGAGWLLSPRSPSGPTDPRRSNIG 120

Qy 121 KVIDTLGCFADLXGYIPVAGPLGGARALAHGVRLVEDGVNATGNLPCGSPSIFLLA 180
 Db 121 KVIDTLGCFADLXGYIPVAGPLGGARALAHGVRLVEDGVNATGNLPCGSPSIFLLA 180

Qy 181 LLSCLTTPASAYVRNVSIGVHVNDGNSNIVTEADVIMHTGCVPCVQEGNSRQW 240
 Db 181 LLSCLTTPASAYVRNVSIGVHVNDGNSNIVTEADVIMHTGCVPCVQEGNSRQW 240

Qy 241 ALPTPLAARNASVPTTIRRHVDLLVGTAAFCASAWYCDLCSIFLYSOLFTEPRRHET 300
 Db 241 ALPTPLAARNASVPTTIRRHVDLLVGTAAFCASAWYCDLCSIFLYSOLFTEPRRHET 300

Qy 301 VQDCNCSITPGHVSCHRMADMMNWSPTTALVYVSQLLRIPOAVDMVAGAHGWGLAGIA 360
 Db 301 VQDCNCSITPGHVSCHRMADMMNWSPTTALVYVSQLLRIPOAVDMVAGAHGWGLAGIA 360

Qy 361 YSMVGWAKULIYVALLPAGVDGETHTTGRVAGHTTSFTSLFSSGASQKQLVNTNGSW 420
 Db 361 YSMVGWAKULIYVALLPAGVDGETHTTGRVAGHTTSFTSLFSSGASQKQLVNTNGSW 420

Qy 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCRIPDPAQGWGPIY--TKPN 478
 Db 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCRIPDPAQGWGPIY--TKPN 478

Qy 479 SSDORPYCWHYAPRCGVVPVPSQVCGPYCYFTPGPVVVGTTDRSGVPTYSGENETVYL 538
 Db 479 SSDORPYCWHYAPRCGVVPVPSQVCGPYCYFTPGPVVVGTTDRSGVPTYSGENETVYL 538

Qy 539 LNNTRPPGWNFGCTWNNSTGFTKTCGPPNCNIGGVGNRT--LICPTCPKPKHPEATYTK 596
 Db 539 LNNTRPPGWNFGCTWNNSTGFTKTCGPPNCNIGGVGNRT--LICPTCPKPKHPEATYTK 596

Qy 541 LNSRTPPGQSWFGCTWNNSTGFTKTCGAPPNCRIEADFNASMDLLCTDCFRKHPDITYIK 600
 Db 541 LNSRTPPGQSWFGCTWNNSTGFTKTCGAPPNCRIEADFNASMDLLCTDCFRKHPDITYIK 600

Qy 597 CGSGFWLTPRLVDYPPYELWHPCTLNFSIFKVRMYGVGZHRNNAACNWRGRCNLED 656
 Db 597 CGSGFWLTPRLVDYPPYELWHPCTLNFSIFKVRMYGVGZHRNNAACNWRGRCNLED 656

Qy 657 RDRSLSPLLSTTEWQILPCAFTLTPALSTGLHLEQNIIVDVQVLYGVGSFAFSFAIK 716
 Db 657 RDRSLSPLLSTTEWQILPCAFTLTPALSTGLHLEQNIIVDVQVLYGVGSFAFSFAIK 716

Qy 717 EYILLFLLADARVACGLMMLIAQBAALENLVVINAASVAGAHGILSFLVFFCAAW 776
 Db 717 EYILLFLLADARVACGLMMLIAQBAALENLVVINAASVAGAHGILSFLVFFCAAW 776

Qy 721 EWWLLFLLADARVACGLMMLILGQBAALAKLVILHAASAAACNGLFVFFVFAAW 780
 Db 721 EWWLLFLLADARVACGLMMLILGQBAALAKLVILHAASAAACNGLFVFFVFAAW 780

Qy 777 YIKGLAPGAAYAFYGVNPLLLLLLLPPPAYALDREMAASCAGAVIAGLVFLTSPYK 836
 Db 777 YIKGLAPGAAYAFYGVNPLLLLLLLPPPAYALDREMAASCAGAVIAGLVFLTSPYK 836

Qy 840 YIKGRVPLATYSLTGLWSFGLLLALPQAYDAYDASHVHGIGIQAALLLVLLTFLTSPYK 840
 Db 840 YIKGRVPLATYSLTGLWSFGLLLALPQAYDAYDASHVHGIGIQAALLLVLLTFLTSPYK 840

Qy 896 VFTRILFWLQYITTRAEAHMQVWVPPNVRGGRDAIILLTCAVHPBELIFDITKLLAIL 896
 Db 896 VFTRILFWLQYITTRAEAHMQVWVPPNVRGGRDAIILLTCAVHPBELIFDITKLLAIL 896

Qy 900 TLLSRFLWLCYLLTLAEAMVQEWAPMQVGRGDGIIWAVAFPCPGVWFDTIKWLLAVL 900
 Db 900 TLLSRFLWLCYLLTLAEAMVQEWAPMQVGRGDGIIWAVAFPCPGVWFDTIKWLLAVL 900

Qy 956 GPLMVIQAGITRPPYFVRAQGLIRACMLRVKAGHYVQWFKLGLTQTYVYNHUTPL 956
 Db 956 GPLMVIQAGITRPPYFVRAQGLIRACMLRVKAGHYVQWFKLGLTQTYVYNHUTPL 956

Qy 960 GPAYLLKAGLTPYFVRAHALLRMCTWRLHAGGRYVQWVLLALGELGTQTYVYDHLTPM 960
 Db 960 GPAYLLKAGLTPYFVRAHALLRMCTWRLHAGGRYVQWVLLALGELGTQTYVYDHLTPM 960

Qy 1016 RDWAHGRDLAVAVPVPVPSAMETKVIITWAGDTPAACDIIILGLPVSARGKIFLGPAD 1016
 Db 1016 RDWAHGRDLAVAVPVPVPSAMETKVIITWAGDTPAACDIIILGLPVSARGKIFLGPAD 1016

Qy 1020 SDWAANGRLDVAVEPIIFSPMEKKVIWGAETAACGDIILHGLPVSARLGRVLLGPAD 1020
 Db 1020 SDWAANGRLDVAVEPIIFSPMEKKVIWGAETAACGDIILHGLPVSARLGRVLLGPAD 1020

Qy 1076 SLEGGWELLAPITAYSQOTRGVLGCIITSITGRDKNOVEGEVQVWSTATQSFATCING 1076
 Db 1076 SLEGGWELLAPITAYSQOTRGVLGCIITSITGRDKNOVEGEVQVWSTATQSFATCING 1076

Qy 1080 GYTSKWSLLAPITAYAQOTRGLLGIIVVSMTRGDKTEQAGEIQVLTSTQSPFGTISG 1080
 Db 1080 GYTSKWSLLAPITAYAQOTRGLLGIIVVSMTRGDKTEQAGEIQVLTSTQSPFGTISG 1080

Qy 1136 VCVTVHAGASKTLGPKGDIITOMYTNVDLIVGWAQPPGARSWTPCSCSSDLVYVTRH 1136
 Db 1136 VCVTVHAGASKTLGPKGDIITOMYTNVDLIVGWAQPPGARSWTPCSCSSDLVYVTRH 1136

Qy 1140 VLWTVHAGNKTLAGSRGVTQWYSSAEGDLVGVWPPPGTKSLPCTCGAVDLYLVRN 1140
 Db 1140 VLWTVHAGNKTLAGSRGVTQWYSSAEGDLVGVWPPPGTKSLPCTCGAVDLYLVRN 1140

Qy 1196 ADVIPVRRGRDGRSGLLSPRFVSYLKSSSGPILLCPSGHVGVFRAAVCTRGVAKAVDFI 1196
 Db 1196 ADVIPVRRGRDGRSGLLSPRFVSYLKSSSGPILLCPSGHVGVFRAAVCTRGVAKAVDFI 1196

Qy 1200 ADVIPARRGRDGRSGLLSPRFVSYLKSSSGPILLCPSGHVGVFRAAVCTRGVAKAVDFI 1200
 Db 1200 ADVIPARRGRDGRSGLLSPRFVSYLKSSSGPILLCPSGHVGVFRAAVCTRGVAKAVDFI 1200

Qy 1256 PVSMETMTRSPFTDNSTPPAVPQTFQVAHLHAPTSGSKSTKVPAAAYAQYKVLVNP 1256
 Db 1256 PVSMETMTRSPFTDNSTPPAVPQTFQVAHLHAPTSGSKSTKVPAAAYAQYKVLVNP 1256

Qy 1260 PVETLIVTRSPFTDNSTPPAVPQTFQVGLHAPTSGSKSTKVPAAAYAQYKVLVNP 1260
 Db 1260 PVETLIVTRSPFTDNSTPPAVPQTFQVGLHAPTSGSKSTKVPAAAYAQYKVLVNP 1260

Qy 1316 SVAATLFGFAGMSKAHGDINIRGTURITITGSIITVSTYKGLADGCGSGGAYDIICD 1316
 Db 1316 SVAATLFGFAGMSKAHGDINIRGTURITITGSIITVSTYKGLADGCGSGGAYDIICD 1316

Qy 1320 SVAATLFGFAGMSKAHGDINIRGTURITITGSIITVSTYKGLADGCGSGGAYDIICD 1320
 Db 1320 SVAATLFGFAGMSKAHGDINIRGTURITITGSIITVSTYKGLADGCGSGGAYDIICD 1320

Qy 1376 ECSTDSSTLIGTIVLDQAEAGARLVATATPPGSVTVPHNIBEIGLSNNGEPTFY 1376
 Db 1376 ECSTDSSTLIGTIVLDQAEAGARLVATATPPGSVTVPHNIBEIGLSNNGEPTFY 1376

Qy 1380 ECHAVDSTLIGTIVLDQAEAGARLVATATPPGSVTVPHNIBEIGLSNNGEPTFY 1380
 Db 1380 ECHAVDSTLIGTIVLDQAEAGARLVATATPPGSVTVPHNIBEIGLSNNGEPTFY 1380

Qy 1436 GKAIPIBAIKGRHLIFCHSKKCKDELAALRGVGLNNAVAYRGLDVSPIPGDVVVVA 1436
 Db 1436 GKAIPIBAIKGRHLIFCHSKKCKDELAALRGVGLNNAVAYRGLDVSPIPGDVVVVA 1436

Qy 1440 GRAIPLSYKGRHLIFCHSKKCKDELAALRGVGLNNAVAYRGLDVSPIPGDVVVVA 1440
 Db 1440 GRAIPLSYKGRHLIFCHSKKCKDELAALRGVGLNNAVAYRGLDVSPIPGDVVVVA 1440

Qy 1496 TDALMTGFTGDPVSVIDNCVTQVDFSLDPTFTTITVTPQDAVSRSORRGRTGGRS 1496
 Db 1496 TDALMTGFTGDPVSVIDNCVTQVDFSLDPTFTTITVTPQDAVSRSORRGRTGGRS 1496

Qy 1500 TDALMTGFTGDPVSVIDNCVTQVDFSLDPTFTTITVTPQDAVSRSORRGRTGGRS 1500
 Db 1500 TDALMTGFTGDPVSVIDNCVTQVDFSLDPTFTTITVTPQDAVSRSORRGRTGGRS 1500

Qy 1556 GIYRFVTPGPRPSGMFDSVYCECYDAGCAWYELTPAETSVRLRAYLNTPEGLPVCQDHL 1556
 Db 1556 GIYRFVTPGPRPSGMFDSVYCECYDAGCAWYELTPAETSVRLRAYLNTPEGLPVCQDHL 1556

Qy 1560 GIYRFVTPGPRPSGMFDSVYCECYDAGCAWYELTPAETSVRLRAYLNTPEGLPVCQDHL 1560
 Db 1560 GIYRFVTPGPRPSGMFDSVYCECYDAGCAWYELTPAETSVRLRAYLNTPEGLPVCQDHL 1560

Qy 1616 PWESVFTGLTHIDAHFLSQTQKQAGNPPYLVAYOATVCARAQAPPPSWDQWKLILKLP 1616
 Db 1616 PWESVFTGLTHIDAHFLSQTQKQAGNPPYLVAYOATVCARAQAPPPSWDQWKLILKLP 1616

Qy 1620 FWEAVFTGLTHIDAHFLSQTQKQAGNPPYLVAYOATVCARAQAPPPSWDQWKLILKLP 1620
 Db 1620 FWEAVFTGLTHIDAHFLSQTQKQAGNPPYLVAYOATVCARAQAPPPSWDQWKLILKLP 1620

Qy 1676 TLHGPTPLLYELGAVONEVILTHPIKYIMACKSADLEVVTSTWLVGGVLAALAAVCLT 1676
 Db 1676 TLHGPTPLLYELGAVONEVILTHPIKYIMACKSADLEVVTSTWLVGGVLAALAAVCLT 1676

Qy 1680 TLVGPTPLLYELGAVONEVILTHPIKYIMACKSADLEVVTSTWLVGGVLAALAAVCLT 1680
 Db 1680 TLVGPTPLLYELGAVONEVILTHPIKYIMACKSADLEVVTSTWLVGGVLAALAAVCLT 1680

Qy 1736 TGSVVIVGRITILSKPAVVPDREVLVYQEFDBMEBCASQLPVEQGMOLAQBFQKALGLL 1736
 Db 1736 TGSVVIVGRITILSKPAVVPDREVLVYQEFDBMEBCASQLPVEQGMOLAQBFQKALGLL 1736

Qy 1740 TGCVCIIIRLHVNOQAVAPDKVLYEAFDBMEBCASRAALIEGORIAEWMKSKIOGLL 1740
 Db 1740 TGCVCIIIRLHVNOQAVAPDKVLYEAFDBMEBCASRAALIEGORIAEWMKSKIOGLL 1740

Qy 1796 QTAATQKQAAAPVAVESKWRALETFWAKHMNPNISIGIYLAGLSTLPGNPAIASMATAS 1796
 Db 1796 QTAATQKQAAAPVAVESKWRALETFWAKHMNPNISIGIYLAGLSTLPGNPAIASMATAS 1796

Qy 1800 QQASKQAOI-QPAVOASWPKVEQFWAKHMNPNISIGIYLAGLSTLPGNPAIASMATAS 1800
 Db 1800 QQASKQAOI-QPAVOASWPKVEQFWAKHMNPNISIGIYLAGLSTLPGNPAIASMATAS 1800

Qy 1856 ITSPLTTONILLFNILGWTAAQLAPSAASAFVAGIAGNAGSICIGIKVLDILAGY 1856
 Db 1856 ITSPLTTONILLFNILGWTAAQLAPSAASAFVAGIAGNAGSICIGIKVLDILAGY 1856

Qy 1860 LTPSLTSTTILLNIGLWLASQIAPPAGATGFVYVGLVGAAGVSGIGLKVLDILAGY 1860
 Db 1860 LTPSLTSTTILLNIGLWLASQIAPPAGATGFVYVGLVGAAGVSGIGLKVLDILAGY 1860

1857 AGVAGALVAFKMGSEVFESTEDLVNLLPAILSPGALVVGVCVCAAILRRHVGGEGAVQWM 1916
1861 AGISGALVARKMGSEFSEMEDVNVLLPGLISPGALVVGVCVCAAILRRHVGGEGAVQWM 1920
1917 NRLIAFASRGNHVSPTHYVBPESDAAARVTOILSSITLQLLKRLHOMNEDCSTPCSGSW 1976
1921 NRLIAFASRGNHVAPTHVYTESDASQRTVLGLSLTISLRLHNMWITDCPTPCSGSW 1980
1977 LRDVMDWICTVLTDKFWLQSKLPRLPGLVFPFLSCQGYKGVWRGDMGIMOTTCGGAQIA 2036
1981 LRDVMDWICTVLTDKFWLQSKLPRLPGLVFPFLSCQGYKGVWRGDMGIMOTTCGGAQIA 2040
2037 GHVXNGSVIRVGPNTCTSNWHTGTPPINAYTTGCTPSPAPNYSRALMRVAEEVYVTRV 2086
2041 GNVRLGSMRITGPTCTMNIWGTTPINCYTGCQCPKPAPEFKIAIWEVAASEVAEYVQH 2100
2097 GDFHYVTGNTDNVCKPCQVAPBFFTEVDGVRHLRYAPACKPRLREDVTQVGLNOYLV 2156
2101 GSYHYITGLTDNLUKVPCLSPSPFFSWDGVQJHRFAPIEKPFPRDEVSFCVGLNSFV 2160
2157 GSQPCPEPDTVLTSMLTDPESHITAEAKRRLARGSPPLASSASQLSAPSLKATCT 2216
2161 GSQPCPEPDTVLTSMLTDPESHITAEAKRRLARGSPPLASSASQLSAPSLKATCT 2220
2217 THDSPPADLLEANLLMRQBMGNITRVESNKVILDSFPLHAEGEREISVAEILR 2276
2221 THGKAYDVMVDANLF---MGSDVTRIESESKVVLDLSDLPVMEERSDLEPSIPSEYML 2276
2277 KSRFPESALPWARDYNNPILLESKDPDYPVPHGCLPPTKAPPIPPRRKXTVLT 2336
2277 PKKFPFPALPWARDYNNPILLESKDPDYPVPHGCLPPTKAPPIPPRRKXTVLT 2336
2337 ESNVSSALAEATKTFG-----SSGSANDSGTATPDLIASDDGDKGSDVESYS 2386
2337 ESSADALQOLAISFGQPPSGDGLSTGDAADSGSTPPDELAL-----SETGSIS 2390
2387 SMPHLEGPPGDDL-----SDGSWSTVSEASBDDVCCMSYTW 2426
2391 SMPLEGGFPDLEPEQVELQPPQGGVTPFGSGSGSWSCSEB-DSDVVCMSYSWT 2449
2427 GALTIPCAAESKLPIPLNSLRLHNMVATTSRSASLRQKVTFTDLQVLDDHYEDV 2486
2450 GALTIPCSFEEKLPIPLNSLRLHNMVATTSRSASLRQKVTFTDLQVLDDHYEDV 2509
2487 LKEMKASTVAKALLSEAEKLTTPHSAKSKGYGAKOVNLSRAVNHRSWVWDL 2546
2510 LKDKLAASKVTARLLTLEAEQTLTPHSAKSKGYGAKOVNLSRAVNHRSWVWDL 2569
2547 EDTETPIDTTIMAKSEVFCVQPKGRKPARLIVFPDGLGVCEKMAIYDVSTLPQAVM 2606
2570 EDTQTEPITINAKNEVFCVDPKGGKAARLIVFPDGLGVCEKMAIYDVSTLPQAVM 2629
2607 GSSYGFQSPQRVEFLVNTWKSCKPMGFYDTRCFDSTVTESDIRVEESYOCDDIAP 2666
2630 GASVGFQSPQRVEFLVNTWKSCKPMGFYDTRCFDSTVTESDIRVEESYOCDDIAP 2689
2667 BARQAIRSLTERLYIGGPLTNSKQNGQYRCRAGSLTSCNTLTCLYKATAACRAK 2726
2690 EARTAHSLTERLYIGGPLTNSKQNGQYRCRAGSLTSCNTLTCLYKATAACRAK 2749
2727 LQDCTMLVNGDVLVICSAGTODAAALRAFTAMTRYSAAPGPPPEYDLEILTS 2786
2750 IIAPTMLVCGDVLVICSAGTODAAALRAFTAMTRYSAAPGPPPEYDLEILTS 2809
2787 SNVSAHDASGRVYLRDPTTPLARAWEATARTPTNSWLGNIIMVAPTILWARMILMT 2846
2810 SNVVALQPGQERRVYLRDPTTPLARAWEATARTPTNSWLGNIIMVAPTILWARMILMT 2869
2847 HFFSILLAQOLEKALDCQIYGACVSIPEPLDQPIIERLHGLSAFTLHSPGHEINRVAS 2906
2870 HFFSILLMAQDITLQNLNFMVYGAUSVSPFLDPLAIIERLHGLDAPSLHTYTPHELTRVAS 2929
2907 CLRKLGVPFLRTWRHARSVRALKLISQGRATTCGRYLFNVAVRTKLTLPAPASQDL 2966

2930 ALRKLGAAPPRAKRSARAVRASLI SRGGRAAVCGRYLFNVAVRTKLTLPPEARLLDL 2989
2967 SGWFVAGYSGDLYHSLSRARPWFPLCLLLLSVGVGVYLLPNR 3010
2990 SSWFTVGAGGDLYHSVRARPRLILLGLLLLFVGVGLFLLPAR 3033
RESULT 8
POLG HCVJ8
ID POLG HCVJ8 STANDARD; PRT; 3033 AA.
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV)
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11115;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kuraai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
homology to reported isolates: comparative study of four distinct
genotypes.";
RL Virology 188:331-341 (1992).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
hydrophobic, suggesting a possible membrane-related function. NS3
and NS5 may play a role in the viral RNA replication.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
precursor polyprotein, commonly with Asp or Glu in the P6
position. Cys or Thr in P1 and Ser or Ala in P1',
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
(RNA) (N).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
lipoprotein envelope. The envelope consists of two proteins:
protein M and glycoprotein E. The nucleocapsid is a complex of
protein C and RNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL: D10988; BAA01761.1; -
PIR: A40250; GNWVJ8.
HSSP: P27958; 1HEI.
MEROPS: S29.001; -
MEROPS: U39.001; -
DR InterPro: IPR009003; Cys Ser trypsin.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR000745; HCV NS4.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRp.

Db 1261 SVAATLGGAYMSKAGHNPNTRTGRTVITGDSITYSTYGRFIADGGCAAGAYIIICD 1320
QY 1317 ECHSTDSITLIGITVLDOAETAGARLVLAATATPPGTVTVEPHNTERIGLSNNGEIPRY 1376
Db 1321 ECHSVDAITLIGITVLDOAETAGARLVLAATATPPGTVTVEPHNTERIGLSNNGEIPRY 1380
QY 1377 GKAIPIEAIKGRHUIFCHSKKCCDELAALTKGLGNNAVYVYRGLEVSVIPPIGDVVVVA 1436
Db 1381 GKAIPIEAIKGRHUIFCHSKKCCDELAALTKGLGNNAVYVYRGLEVSVIPPIGDVVVVA 1440
QY 1437 TDALMTGTGDPDSVIDCNTCTVTVDSLDPTFLEITTVQDVAVSQRGRTRGRS 1496
Db 1441 TDALMTGTGDPDSVIDCNVAVSQVDFSLDPTFTTQTVQDVAVSQRGRTRGRS 1500
QY 1497 GIYREVTGPERSGMFDSSVLCEDCAGCAWYELTPAETSVSLRAYLNTGELPVQCDHLE 1556
Db 1501 GYRVYSSGERSGMDSSVLCEDCAGCAWYELTPAETSVSLRAYLNTGELPVQCDHLE 1560
QY 1557 FHSYFTGLTHIDAHFLSOTKQAGNFFVLYAYQATVCARAQAPPSMDQWKKLIRLKP 1616
Db 1561 FMEAVFTGLTHIDAHFLSOTKQAGNFFVLYAYQATVCARAQAPPSMDQWKKLIRLKP 1620
QY 1617 TLHGPTPLLYRLGAVQNEVLTHPIKXIMACWSADLEVVSTWVVGVLAAALAYCLT 1676
Db 1621 TLHGPTPLLYRLGAVQNEVLTHPIKXIMACWSADLEVVSTWVVGVLAAALAYCLT 1680
QY 1677 TGSVVVGRILISGRPAVVDREVLVYQFDEMEECASQLPYIEQGMQLAEPQKXALGLL 1736
Db 1681 TGCISIGRLNDRVVVAPDKEILVYAEFDEMEECASKAALIEBQORVAEMLKSKIQLL 1740
QY 1737 QTATQAAAPAVVESKWEALBETFAKMKWNFIISGIOYLAGLSTLPGNPALASMAFTAS 1796
Db 1741 QOATQOADIOPAIOSKPEKLEQFAKMKWNFIISGIOYLAGLSTLPGNPALASMAFTAS 1800
QY 1797 ITSPLTQNTLLFNILGWWAAQLAPPSAASAFVCGAGIAGAAVGSIGLKVLDITLAVG 1856
Db 1801 LTSPLTQNTLLFNILGWWAAQLAPPSAASAFVCGAGIAGAAVGSIGLKVLDITLAVG 1860
QY 1857 AGVAGALVAFKMSGEVSTEDLVNLLPAILSPGALVGVVCAALIRRHVGPGEVAVQVM 1916
Db 1861 AGVAGALVAFKMSGEVSTEDLVNLLPAILSPGALVGVVCAALIRRHVGPGEVAVQVM 1920
QY 1917 NLLTAFAGRGHNVSTHVPVSDAARVTOILSLTITOLKRLHOWNEDCSTPCSGW 1976
Db 1921 NLLTAFAGRGHNVSTHVPVSDAARVTOILSLTITOLKRLHOWNEDCSTPCSGW 1980
QY 1977 LRVDWDITCTVLTDPKTLQKLLPRLPGVPLSCORGKGVWRGDGIMGTCTPCGAGIA 2036
Db 1981 LODIWDVCSILTDKFWLSSKLLPMPGIPFISCKYKGVWAGTGVWTCPCGANIS 2040
QY 2037 GHVKNMGRIVGRPTCSNTHGTTPPINAYTGTCTPSPAPNYSRALMRVAABEYVEVTRV 2096
Db 2041 GHVRMGTKITGPKTCLNLQCTTPINCYTEGPCVPKPPNPKYKTAIRWVAASEYVEVTH 2100
QY 2097 GDEYVTCMTDNVCKPCQVAPFEFTVGVRLHRYAPACKLLREDTVQVGLNQVLY 2156
Db 2101 GSFVTVGLTDNLKVPCCQVAPFEFTVGVRLHRYAPACKLLREDTVQVGLNQVLY 2160
QY 2157 GSQPCPEPDPVTLTSLMTLTPSHITAEAKRLIARGSPPLSAGSSASQLSAPSLKATCT 2216
Db 2161 GSQPCPEPDPVTLTSLMTLTPSHITAEAKRLIARGSPPLSAGSSASQLSAPSLKATCT 2220
QY 2217 THHDSPDADLEALNLLWQEQNGNITRVESNKNVILDSFPLHAEGDEREISVAEILR 2276
Db 2221 THKTAYDCMDVANDLF----MGGDVTRIESDKVILDSLSMTVEEDRPSVSEVLY 2276
QY 2277 KSRKFPESALPWARDNPNPILLESKDPYVPPVHGCPLPPTKAPPIPPPRKRTVLT 2336
Db 2277 KRRKFPESALPWARDNPNPILLESKDPYVPPVHGCPLPPTKAPPIPPPRKRTVLT 2336
QY 2337 ESNVSSALAEALATKTFG-----SSGSSAVDSGTATALPDLASDGGK--GSDVESYSS 2387

Db 2337 QDNVEGLREMADKVLSPLQDNNDSGHSTGADTG-----GDIVQOPSDETAASEAGLS 2391
QY 2388 MPELEGEPGDDI-----SDG-----SNSTVSEASSEDVUCCMSYTWG 2427
Db 2392 MPELEGEPGDDI-----SDG-----SNSTVSEASSEDVUCCMSYTWG 2450
QY 2428 ALITPCAAEESKLPINLSNLSLRHNMVYATTSRSASLRKOKKVTDFRLOVLDDHYRDL 2487
Db 2451 ALITPCPEEBEKLPIPLSNLSMRHNMVYATTSRSASLRKOKKVTDFRLOVLDDHYRDL 2510
QY 2488 KEMKAKASTVKAULSTEEACKLTPHSAKSKFYGAKOVNLSRAVNHRSWEDLLE 2547
Db 2511 QCVRAAASKVARILITVEEACALTPHSAKSKFYGAKOVNLSRAVNHRSWEDLLE 2570
QY 2548 DTEPFDITIMAKSEVFCVQPEKGRKPARLIVPDLGVRVCEKMAVYDVVSTLPOAVMG 2607
Db 2571 DQTFPDIITIMAKNEVFCIDPTKGGKPARLIVPDLGVRVCEKMAVYDVVSTLPOAVMG 2630
QY 2608 SSGFYQSPKQORVEFLVNTWKSCKPMGFSDYTRCFDSTVTESDIRVEESIYQCCDLAPE 2667
Db 2631 PSYFYQSPKQORVEFLVNTWKSCKPMGFSDYTRCFDSTVTESDIRVEESIYQCCDLAPE 2690
QY 2668 ARQAIRSITELRYTGGPLTNSKQNGCYRRCRAGSVLTTSCGNTLTCTYLKATAACRAKL 2727
Db 2691 ARTVHSLTETRYTGGPLTNSKQNGCYRRCRAGSVLTTSCGNTLTCTYLKATAACRAKL 2750
QY 2728 QDCTMLVNGDDLVLVICSAGTOEDAAALRAPTAMTRYSAPPDPPOPEYDLELITSCS 2787
Db 2751 VDFVNLVCGDDLVLVICSAGTOEDAAALRAPTAMTRYSAPPDPPOPEYDLELITSCS 2810
QY 2788 NVSVAHASGRVYLLTRDPTPLARAAMETARTPTINSLGNIIMYAPTILWARMILMT 2847
Db 2811 NVSVALSRRGRYRFLTRDPTPLARAAMETARTPTINSLGNIIMYAPTILWARMILMT 2870
QY 2848 FFSILLAQEOLKALDCQIYACACYSIEPLDLPQILERLHGLSAPTHSYSPGEINRVASC 2907
Db 2871 FFSILLAQEOLKALDCQIYACACYSIEPLDLPQILERLHGLSAPTHSYSPGEINRVASC 2930
QY 2908 LRKLGVEPLRTWRERARSVRKLLSQGGRATCGRYLFWAVRTKLTIPPAASQLDLS 2967
Db 2931 LRKLGVEPLRTWRERARSVRKLLSQGGRATCGRYLFWAVRTKLTIPPAASQLDLS 2990
QY 2968 GWFVAGVSGGDIYHSLRARPRWPPCLLLSVGVGVIYLLPNR 3010
Db 2991 GWFVAGVSGGDIYHSLRARPRWPPCLLLSVGVGVIYLLPNR 3033

RESULT 9
POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix protein (Envelope protein M); Major envelope protein E; Nonstructural proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J5) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11112;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RL genotypes." Virol. 188:331-341 (1992).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.

CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins,
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10075; BAA00969.1; --
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
KW INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;
Query Match
Best Local Similarity 19.4%; Score 3101; DB 1; Length 737;
Matches 533; Conservative 87; Mismatches 113; Indels 4; Gaps 2;
QY 1 MSTNPKPQRTKNTNRRPQVKGQIVGGVYLLPRGRPLGVATRKASRSQPRG 60
DB 1 MSTNPKPQRTKNTNRRPQVKGQIVGGVYLLPRGRPLGVATRKASRSQPRG 60
QY 61 RQPIKARPEGRAPQGPWPLVYNGEIGWAGWLLSRGSPSPSGPTDPRRSNLG 120
DB 61 RQPIKARPEGRAPQGPWPLVYNGEIGWAGWLLSRGSPSPSGPTDPRRSNLG 120
QY 121 KVIDTLCTGADLMGYIFLVGAPLGGARALAHGVRLVDGWNVATGNLPGCSFIFLLA 180
DB 121 KVIDTLCTGADLMGYIFLVGAPLGGARALAHGVRLVDGWNVATGNLPGCSFIFLLA 180
QY 181 LLSCLTTPAGAYEVNVSIGYHVNTDCNSGSIIVTEADVIMETPGCVQVQEGNSRCW 240
DB 181 LLSCLTTPAGAYEVNVSIGYHVNTDCNSGSIIVTEADVIMETPGCVQVQEGNSRCW 240
QY 241 ALTYTLAARNASVPTTIRHVDLVLTGAFCASMYGDLGCSFIYNSQLTFTSPRRHET 300
DB 241 ALTYTLAARNASVPTTIRHVDLVLTGAFCASMYGDLGCSFIYNSQLTFTSPRRHET 300

DB 241 PVSPNVAVRQPCALTQGLRTHIDMVVVSATLCSALYVGDLCGGVMLAAQMFIVSPQHWF 300
QY 301 VQDCNCSIYPCHVSGHMANWMMNSPTTALVVSQLLRIPQAVDVMVAGHWGLAGLA 360
DB 301 VQDCNCSIYPCHVSGHMANWMMNSPTTATMILAYAMRVDEVIDIISGAHWGMFGLA 360
QY 361 YVSMVGNWAKVLLIALLFAGVDGTHHTGTVAGHTTSGFTSLPSSGASQKTQVNTNGSW 420
DB 361 YFSMOGAWAKVUVILLAAAGVDANTRVAGSAATTRGFTSMFSSGSKQLQLINTNGSW 420
QY 421 HINTALNDSLOTGFFAALFYAHKENSNGCERMAACRPIDWFAQGWGPITY--TKPN 478
DB 421 HINTALNDSLNTGFIASLFYVNRFNSSGCPHRLSVCRSIEAFRIGWGLQVNDVNT 480
QY 479 SSDORPYCHYAPRCQGVFASQVCGPYVCTPSVAVVGTTRDSGVITYSWGENETVML 538
DB 481 PEDMRPYCHYPPKCGIGVFAERSVCGPYVCTPSVAVVGTTRDARGVPTYTWGENETDVL 540
QY 539 LNTTRPPQGNWFGCTWMNSTGFTKTCGGPCNTGGVGNRT--LICPTDCPKHPEATYTK 596
DB 541 LNSTPRPGSGWFGCTWMNSTGFTKTCGAPPCIRPADFNASTDLCTDCPKHSDATYIK 600
QY 597 CGSGPMLTPCLVDYPRYLMEYPTCLNFISIFKVMYVGGVEHRLNAACNTRGRCNLED 656
DB 601 CGSGPMLTPCMVDYPRYLMEYPTVMYSIFKRMVYVGGVEHRLTAACNTRGRCNLED 660
QY 657 RDRSELSPLLSTEWIILCATITLPAALSTGLHLHONIVDQVLYGVCSAFVSPFAKW 716
DB 661 RDRSQSPLHSTETWAILPCTYSDLPALSTGLHLHONIVDQVLYGVCSAFVSPFAKW 720
QY 717 EYIILFELLADARVCA 733
DB 721 EYVLLFELLADARVCA 737
RESULT 10
POLG HCVJ7
ID POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC 27961;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE protein [Envelope protein M]; Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda S., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; D10077; BAA00971.1; --
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186662; HCV NS1; 1.
 KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
 Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122E2 CRC64;
 Query Match 19.3%; Score 3093; DB 1: Length 737;
 Best Local Similarity 72.5%; Pred. No. 1.4e-185;
 Matches 534; Conservative 80; Mismatches 119; Indels 4; Gaps 2;
 QY 1 MSTNPKQRTKXNTNRPPQDVKPPGGQIVGGVYLLPRRGLGVATRKASERSQPRG 60
 DB 1 MSTNPKQRTKXNTNRPPQDVKPPGGQIVGGVYLLPRRGLGVATRKASERSQPRG 60
 QY 61 RROPIKARRPEGRMAQCPYWPVLYGNEGLHAGWLLSPRGRPSWGPTDPRRSNRLG 120
 DB 61 RROPIKARRPEGRMAQCPYWPVLYGNEGLHAGWLLSPRGRPSWGPTDPRRSNRLG 120
 QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAAALAHGVRVLEQGVNATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPLVGAFLGGAAALAHGVRVLEQGVNATGNLPGCSFSIFLLA 180
 QY 181 LLSCLHIPSAYEVNVSIGHYHVNDCSNSSIVYEADVMHTPGCVPCQEGNSRCRW 240
 DB 181 LLSCLHIPSAYEVNVSIGHYHVNDCSNSSIVYEADVMHTPGCVPCQEGNSRCRW 240
 QY 241 ALPTPLAARNASVPTTIRRHVLLVGTAAFCASAMVYDLCGSIPLVSLQFTSPRHET 300
 DB 241 ALPTPLAARNASVPTTIRRHVLLVGTAAFCASAMVYDLCGSIPLVSLQFTSPRHET 300
 QY 301 VQDNCNSIYPGHVSGRHMAWMMNNSPTTALVVSQLLRIPQAVDVMAGHVGVLG 360
 DB 301 VQDNCNSIYPGHVSGRHMAWMMNNSPTTALVVSQLLRIPQAVDVMAGHVGVLG 360
 QY 361 YYSWGVNKAIVALLFAGVDCGTHGTGVAHTSGFTSLFSSGASQKIQVNTNGSW 420
 DB 361 YYSWGVNKAIVALLFAGVDCGTHGTGVAHTSGFTSLFSSGASQKIQVNTNGSW 420
 QY 421 HINRTALNCNDSLQTFGAALFYVRNNSGGERLSCEKLDLDFRIGWGLTETVNTN 480

DB 421 HINRTALNCNDSLQTFGAALFYVRNNSGGERLSCEKLDLDFRIGWGLTETVNTN 480
 QY 479 SSQRPYCHYIAPRCPGVVPASQVCGVYCTFTSPVVGTTDRSGVPTYSWGNETDVML 538
 DB 481 EEDMRPYCHYIAPRCPGVVPASQVCGVYCTFTSPVVGTTDRSGVPTYSWGNETDVML 540
 QY 539 LNNTRPEQGNWFCGTWNNSTGFTKTCGGPQNTGGVGNRT--LTCPTDCERKHEATYTK 596
 DB 541 LNNTRPEQGNWFCGTWNNSTGFTKTCGGPQNTGGVGNRT--LTCPTDCERKHEATYTK 600
 QY 597 CGSGPMLTPRCLVDYFYRLMWHYPTLNFSTFKVMYVGGVEHRLNAAACNWTGRNCNLED 656
 DB 601 CGSGPMLTPRCLVDYFYRLMWHYPTLNFSTFKVMYVGGVEHRLNAAACNWTGRNCNLED 660
 QY 657 RDRSELSPILLSTTEWQIPLCAFTTTPALSTGLIHONIVDQVLYGVGSAPVSPAIKW 716
 DB 661 RDRSELSPILLSTTEWQIPLCAFTTTPALSTGLIHONIVDQVLYGVGSAPVSPAIKW 720
 QY 717 EYILLFLLLADARVCA 733
 DB 721 EYILLFLLLADARVCA 737
 RESULT 11
 ID POLG_HCVJ2 STANDARD; PRT; 513 AA.
 AC P27959;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus (isolate HC-J2) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11111;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Iizuka H.,
 Tanaka T., Fukuda S., Tsuda F., Mishihiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes".
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; D10077; BAA00968.1; --
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186662; HCV NS1; 1.


```

RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53134; CAA37294.1; -.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01542; HCV env; 1.
DR Pfam; PF01539; HCV core; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 75 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 76 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 321 321
SQ SEQUENCE 321 AA; 34074 MW; B2EB83F521C3B520 CRC64;

Query Match
Best Local Similarity 9.1%; Score 1453; DB 1; Length 321;
Matches 262; Conservative 28; Mismatches 31; Indels 0; Gaps 0;

QY 117 RNLGKVIDTLTCGFADLMGYIPLVGAAPLGGAARALAHGVRLVDGVNATGNLPGCSFSI 176
Dd 1 RNLGKVIDTLTCGFADLMGYIPLVGAAPLGGAARALAHGVRLVDGVNATGNLPGCSFSI 60

QY 177 FLALLSCLTIPASAYEVRNVSGIYHVNDSCNSIIVVEADVIHMTGCVPCVQEGNSS 236
Dd 61 FLALLSCLTIPASAYEVRNVSGIYHVNDSCNSIIVVEADVIHMTGCVPCVQEGNSS 120

QY 237 RCWVALTPTLARNASVPTTIRRHVDLLVGAAPFCASMYGDLGSIPLVSQLFTFSR 296
Dd 121 RCWVAATPTVATRDGLPTTQRLRHIDLLVGSATLCSALYVGDLCGSIPLVSQLFTFSR 180

QY 297 RHETVQDCNSIYPGHVSGHRVAMNMNWSPTTALVYSQRLRIPOAVYDMVAGAHGVL 356
Dd 181 RHWTQGCNSIYPGHITGHRVAMNMNWSPTTALVVAQLLRIPOAILDIAGAHGVL 240

QY 357 AGLAYVSGVGNWAKVLIIVALLFAGYDGHITTCRVAGHTTSGFTSLFSGASQKIQNT 416
Dd 241 AGLAYVSGVGNWAKVLIIVALLFAGYDAETVTGGSAAHGALIASLFGQARQNIQLNT 300

QY 417 NGSWHINRTALNCNDSLOTGF 437
Dd 301 NGSWHINRTALNCNDSINTGW 321

```

```

RESULT 15
ID POLG_HCVH8 STANDARD; PRT; 321 AA.
AC P27956;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (fragment).
OS Hepatitis C virus (isolate HCT18) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11110;
RN [1]_TaxID=11110;
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53131; CAA37291.1; -.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 76 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 77 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 321 321
SQ SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;

Query Match
Best Local Similarity 9.0%; Score 1441; DB 1; Length 321;
Matches 257; Conservative 32; Mismatches 32; Indels 0; Gaps 0;

QY 117 RNLGKVIDTLTCGFADLMGYIPLVGAAPLGGAARALAHGVRLVDGVNATGNLPGCSFSI 176
Dd 1 RNLGKVIDTLTCGFADLMGYIPLVGAAPLGGAARALAHGVRLVDGVNATGNLPGCSFSI 60

QY 177 FLALLSCLTIPASAYEVRNVSGIYHVNDSCNSIIVVEADVIHMTGCVPCVQEGNSS 236
Dd 61 FLALLSCLTIPASAHQVRNVSTGLYHVNDPCNSIIVVEADAILHTPGCVPCVQEGNV 120

```

```

QY 237 RCWVALPTTLAARNASVPTTIRRHVDLLVGTAAFCSAMVYVGDLCGSIFLVSQLFTEPR 296
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
121 RCWVAVTPTVATRGKLPPTQLRRHIDLIVGSATLCSALYVGDLCGSVELVQLFTEPR 180
QY 297 RHETVQDCNCSIYEGHVSCHRMAMDMNNSPTTALVVSQLLRIPQAVVDMVAGAHGVL 356
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
181 RHWTQGCNCSIYEGHITGHRMAMDMNNSPTTALVNAQLLRIPQAIMDMVAGAHGVL 240
QY 357 AGLAYYSVMVGNWAKVLIVLIFAGVDGETHITIGRVAGHTTSGFTSLFSSGASQIOLVNT 416
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
241 AGIAYFSMVGNWAKVLVLLVLLFAGVDAETYTSGGNAGHTMTGIVRFPAPGPKQNVHLINT 300
QY 417 NGSWHINRTALNCNDSLQTF 437
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
301 NGSWHINRTALNCNDSLNTGW 321

```

Search completed: April 8, 2004, 10:47:07
Job time : 36 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2004, 10:43:42 ; Search time 39 Seconds
(without alignments)
7424.013 Million cell updates/sec

Title: US-09-662-454-3

Perfect score: 16009

Sequence: 1 MSTNPKPQKTKRNTNRRPQ.....PFLCLLLSVGVGYLLPNR 3010

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15367	96.0	3010	1 GNVVCJ	genome polypotein
2	15253	95.3	3010	1 GNVVTC	genome polypotein
3	15236	95.2	3010	1 A45573	genome polypotein
4	15062	94.1	3010	1 GNVVTV	genome polypotein
5	15052	94.0	3010	1 GNVVTC	genome polypotein
6	14119.5	88.2	3011	1 GNVVTC	genome polypotein
7	14008.5	87.5	3011	1 S40770	genome polypotein
8	13935.5	87.0	3011	1 GNVVCH	genome polypotein
9	12496	78.1	3014	1 JCS620	genome polypotein
10	12036.5	75.2	3033	1 JQ1303	genome polypotein
11	11972.5	74.8	3033	1 GNVVJ8	genome polypotein
12	3977	24.8	782	2 S18031	genome polypotein
13	3952	24.7	782	2 S19876	genome polypotein
14	3942	24.6	782	2 S18032	genome polypotein
15	3922	24.5	782	2 S19875	genome polypotein
16	3916	24.5	782	2 PN3677	hypothetical prote
17	3473	21.7	876	2 PC2219	polypeptide - hepa
18	3051	19.1	640	2 JQ1584	genome polypotein
19	3023	18.9	716	2 JQ1366	polypotein - hepa
20	2674	16.7	513	2 PC1284	genome polypotein
21	2581.5	16.1	523	2 JQ1926	polypotein - hepa
22	2577.5	16.1	520	2 JQ1925	polypotein - hepa
23	2513	15.7	550	2 JH0711	genome polypotein
24	2426	15.2	874	2 JQ0893	genome polypotein
25	2417	15.1	874	2 JQ0881	genome polypotein
26	2393	14.9	513	2 A44150	structural protein
27	2387	14.9	492	2 PS0326	polypotein - hepa
28	2286.5	14.3	3005	2 C08841	polypotein - dour
29	2251	14.1	441	2 S12707	genome polypotein

ALIGNMENTS

RESULT 1

GNVVCJ

genome polypotein - hepatitis C virus (strain J)

N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001

C:Accession: A39253; PS0086

R:Kato, N.; Hijikata, M.; Ohtsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto

Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients

A:Reference number: A39253; PMID:91088550; PMID:2175903

A:Accession: A39253

A:Molecule type: genomic RNA

A:Residues: 1-3010 <KAT>

A:Cross-references: GB:D30208; NID:g221610; PIDN:EAA14233.1; PID:g221611

R:Kato, N.; Ohkoshi, S.; Shimotohno, K.

Proc. Jpn. Acad. 65B, 219-223, 1989

A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari

A:Reference number: PS0085

A:Accession: PS0086

A:Molecule type: genomic RNA

A:Residues: 2650-2707 <KA2>

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polypeptide have not been determined.

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serir

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <32M>

F:192-389/Product: major envelope protein E #status predicted <MSE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polypeptide

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: DRX motif

F:1316-1862/Product: nonstructural protein NS4a #status predicted <NA4>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NA4B>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,230,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 95.0%; Score 15367; DB 1; Length 3010;

Best Local Similarity 95.2%; Pred. No. 0;

Matches 2866; Conservative 64; Mismatches 80; Indels 0; Gaps 0;

Qv 1 MSTNPKPQKTKRNTNRRPQDKPFGGQIVGGVYLLPRRPRGLGVATRKASRSQPRG 60

Db 1 MSTNPKPQKTKRNTNRRPQDKPFGGQIVGGVYLLPRRPRGLGVATRKATRSQPRG 60

Qv 61 RRQIPKARRREGAWAQPGYEWPLYGNEGAGWLLSPFGSRFSWGPTDPRRSRLG 120

Db 61 RRQIPKARRREGAWAQPGYEWPLYGNEGAGWLLSPFGSRFSWGPTDPRRSRLG 120

QY 121 KVIDTLTCGPADLMGYIPLVGCAPLGGAARALAHGVRLDGVNATGNLPCSCPSIFLLA 180
 Db 121 KVIDTLTCGPADLMGYIPLVGCAPLGGAARALAHGVRLDGVNATGNLPCSCPSIFLLA 180
 QY 181 LLSCLITIPASAYEVNVSIGIYHVINDCSNSSIIVVEAADVIMETPCVPCVCEGSSRCWV 240
 Db 181 LLSCLITIPASAYEVNVSIGIYHVINDCSNSSIIVVEAADVIMETPCVPCVCEGSSRCWV 240
 QY 241 ALTPTLAARNASVTTTIRRHVDLLVGTAAFCASMYVGDLCGSIPLVSQLTFFSPRHET 300
 Db 241 ALTPTLAARNASVTTTIRRHVDLLVGTAAFCASMYVGDLCGSIPLVSQLTFFSPRHET 300
 QY 301 VQDCNCSITPCYCHSHRMWMMWNSPTTALVVSQLLRIPOAVVDMVAGAHWGLAGLA 360
 Db 301 VQDCNCSITPCYCHSHRMWMMWNSPTTALVVSQLLRIPOAVVDMVAGAHWGLAGLA 360
 QY 361 YYSVMGNWAKYLIIVALLFAGYDGETTTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420
 Db 361 YYSVMGNWAKYLIIVALLFAGYDGETTTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420
 QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSCPCERMASCRPIDWFAQGWGITTKNSS 480
 Db 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSCPCERMASCRPIDWFAQGWGITTKNSS 480
 QY 481 DORPYCHYAPRCGVVPASOVCGPVYCFPTSPVVVGTDRSGVPTYSWGENETDVLN 540
 Db 481 DORPYCHYAPRCGVVPASOVCGPVYCFPTSPVVVGTDRSGVPTYSWGENETDVLN 540
 QY 541 NTRPPQGNWGCCTWNNSTGFTKICGGPPCNIGGVNRTLICPTDCFRKHPEATYKCGSG 600
 Db 541 NTRPPQGNWGCCTWNNSTGFTKICGGPPCNIGGVNRTLICPTDCFRKHPEATYKCGSG 600
 QY 601 PMLAPRCVADYEVYLVHYPCTNLSIFKVRMYGVGVHRIINAAACNWRGECNLEDRDRS 660
 Db 601 PMLAPRCVADYEVYLVHYPCTNLSIFKVRMYGVGVHRIINAAACNWRGECNLEDRDRS 660
 QY 661 ELSPLLLSTTEWQILPCLPAFTLTPALSTGLIHLHOKIVDVQVLYGVGSFVFAIKWEYIL 720
 Db 661 ELSPLLLSTTEWQILPCLPAFTLTPALSTGLIHLHOKIVDVQVLYGVGSFVFAIKWEYIL 720
 QY 721 LLFLLADARVACALWMLLIIAQAEPAALENIIVLNAASVAGAHGILSLFVFFCAAWYIK 780
 Db 721 LLFLLADARVACALWMLLIIAQAEPAALENIIVLNAASVAGAHGILSLFVFFCAAWYIK 780
 QY 781 RLAPCAAYFVGVVPLLLILALPRAYALDREMAASCGLVLYGLVPLT-SPYKXVPLT 840
 Db 781 RLAPCAAYFVGVVPLLLILALPRAYALDREMAASCGLVLYGLVPLT-SPYKXVPLT 840
 QY 841 RLIMWLQYFITRAEAHMVQVPPPLNVGRGRDAIILLTCAVHPELIFDITKLLAILGLPLM 900
 Db 841 RLIMWLQYFITRAEAHMVQVPPPLNVGRGRDAIILLTCAVHPELIFDITKLLAILGLPLM 900
 QY 901 VLQAGITRVPYFVRAQGLIRACMLVRKAGHYQVFMKLGALTGTIVYVHLLPLRDWA 960
 Db 901 VLQAGITRVPYFVRAQGLIRACMLVRKAGHYQVFMKLGALTGTIVYVHLLPLRDWA 960
 QY 961 HAGLEDLAVAVEPVVFSAMETKVIITWAGDTAACGDIILGLPVSARGKEIIFLGPADSLEG 1020
 Db 961 HAGLEDLAVAVEPVVFSAMETKVIITWAGDTAACGDIILGLPVSARGKEIIFLGPADSLEG 1020
 QY 1021 QGWRLLAPITAYSQOTRGVLGCIITSLTGRDNQVGEVQVVSATQSFATCINGVCWT 1080
 Db 1021 QGWRLLAPITAYSQOTRGVLGCIITSLTGRDNQVGEVQVVSATQSFATCINGVCWT 1080
 QY 1081 VYHGAGSKTLAPGKPIITQMTNVNDLIVGWQAPEGARSMTPCSCGSDLLIVTRHADVI 1140
 Db 1081 VYHGAGSKTLAPGKPIITQMTNVNDLIVGWQAPEGARSMTPCSCGSDLLIVTRHADVI 1140
 QY 1141 PVRRGDSRGILLSRPVSYLYKSSGGPILLCPSGHWGVGFRAAVCTRGVAXAVDFIPVES 1200
 Db 1141 PVRRGDSRGILLSRPVSYLYKSSGGPILLCPSGHWGVGFRAAVCTRGVAXAVDFIPVES 1200

QY 1201 KETMTSPVETDNSTPAPVQOTFOVAHLHAPTQSGKSTKVPAAVAAQGVKVLVNPVAA 1260
 Db 1201 KETMTSPVETDNSTPAPVQOTFOVAHLHAPTQSGKSTKVPAAVAAQGVKVLVNPVAA 1260
 QY 1261 TLGFGAYMSKAHIGIDNIRITGVATITITGGISITYTCKFLADGGCSGAYDIILICECHS 1320
 Db 1261 TLGFGAYMSKAHIGIDNIRITGVATITITGGISITYTCKFLADGGCSGAYDIILICECHS 1320
 QY 1321 TDSFTTILGIGTVDLDOAETAGARLVVATATPQSGVTVVPHNIEEIGLSNNGETPFYKAI 1380
 Db 1321 TDSFTTILGIGTVDLDOAETAGARLVVATATPQSGVTVVPHNIEEIGLSNNGETPFYKAI 1380
 QY 1381 PIFAIKGRHLIFCHSKKCDLAAKLITGLINAVAYYRGLDVSVIPIGDVVVAATDAL 1440
 Db 1381 PIFAIKGRHLIFCHSKKCDLAAKLITGLINAVAYYRGLDVSVIPIGDVVVAATDAL 1440
 QY 1441 MTGFTGDFDVIDCNTCVTVTVDFSDPTFTTITTTVPQDAVRSQRRTGRTGRSGIYR 1500
 Db 1441 MTGFTGDFDVIDCNTCVTVTVDFSDPTFTTITTTVPQDAVRSQRRTGRTGRSGIYR 1500
 QY 1501 FVTPGSRPSCMPDSSVLCBCEYDAGCAWYELTPAETSVRLRAYLNTPLPVCODHLEFWES 1560
 Db 1501 FVTPGSRPSCMPDSSVLCBCEYDAGCAWYELTPAETSVRLRAYLNTPLPVCODHLEFWES 1560
 QY 1561 VFTGLTHIDAHFTSQTQAGDNFPYLVAQATVCAZAAQAPPPSWDQWKKLIRLKTPLHG 1620
 Db 1561 VFTGLTHIDAHFTSQTQAGDNFPYLVAQATVCAZAAQAPPPSWDQWKKLIRLKTPLHG 1620
 QY 1621 PTPLLVRLGAVQNEVLTHPIITKYIMACMSADLEVVTSTWLVGGVLAALAAVCLTGTGSV 1680
 Db 1621 PTPLLVRLGAVQNEVLTHPIITKYIMACMSADLEVVTSTWLVGGVLAALAAVCLTGTGSV 1680
 QY 1681 VIYGRILLGKPAVDPREVLYOEFDEMBCASOLPIEOMOLABQFKOKALGLLQAT 1740
 Db 1681 VIYGRILLGKPAVDPREVLYOEFDEMBCASOLPIEOMOLABQFKOKALGLLQAT 1740
 QY 1741 KQEAAPVSVESKWRALETFWAKHMNFISGIIQYLAGSLTLPNPAIASLMAFTASITSP 1800
 Db 1741 KQEAAPVSVESKWRALETFWAKHMNFISGIIQYLAGSLTLPNPAIASLMAFTASITSP 1800
 QY 1801 LTTQNTLLFNILGWNAAQAPPASAFVAGIAGAAVSGISIGLKVILVILLAGYAGVA 1860
 Db 1801 LTTQNTLLFNILGWNAAQAPPASAFVAGIAGAAVSGISIGLKVILVILLAGYAGVA 1860
 QY 1861 GALVAFKVMGEVPSSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEVQVWMAKLI 1920
 Db 1861 GALVAFKVMGEVPSSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEVQVWMAKLI 1920
 QY 1921 AFASRGNHVSPTHYVVPESDAAARVTQILSSITITQLKRLHQWINEJCSPTPCSGMLRDV 1980
 Db 1921 AFASRGNHVSPTHYVVPESDAAARVTQILSSITITQLKRLHQWINEJCSPTPCSGMLRDV 1980
 QY 1981 WDMCTVLTDPKWLQSKLLPRLPGVPEFLSCORGXKGVWGDGIMOTTCBGQIAGHV 2040
 Db 1981 WDMCTVLTDPKWLQSKLLPRLPGVPEFLSCORGXKGVWGDGIMOTTCBGQIAGHV 2040
 QY 2041 NGSMRIIVGPKTCSNTHGTFPIINAYITGPCTPAPNYSALWVAEVEVTRVGDHF 2100
 Db 2041 NGSMRIIVGPKTCSNTHGTFPIINAYITGPCTPAPNYSALWVAEVEVTRVGDHF 2100
 QY 2101 YVTGMTTDNVKPCQVPAPEFFTEVDGVRHRYAPACKPLLRREDVTFQVGLNQYVGSOL 2160
 Db 2101 YVTGMTTDNVKPCQVPAPEFFTEVDGVRHRYAPACKPLLRREDVTFQVGLNQYVGSOL 2160
 QY 2161 PCPEPDPVTLTSMITDPSHITATKARLARGSPSPSLASSASQLSAPLKATCTTHD 2220
 Db 2161 PCPEPDPVTLTSMITDPSHITATKARLARGSPSPSLASSASQLSAPLKATCTTHD 2220
 QY 2221 SPADLLEALLWEOQMGMITRVESENKVVILDSPEPLHAGCERBEISVAAEILKRSK 2280
 Db 2221 SPADLLEALLWEOQMGMITRVESENKVVILDSPEPLHAGCERBEISVAAEILKRSK 2280
 QY 2281 FPSALPFWASPDVNPPLLESWKDPDYVPPVHVHGCPLPPTKAPPIPPPRKRTVWLTESNV 2340

```
Db 2281 FPPALPIWAPDYNPELLESWKPDPYVPVWVHGCPLPSKAPPIPPRRKRTVLTSTV 2340
Qy 2341 SSALAEATKTTGSGSSAVDSGTATAPDLASDDGKSDVESYSSMPLEGEFGDPL 2400
Db 2341 SSALAEATKTTGSGSSAVDSGTATGPPDQASDDGKSDVESYSSMPLEGEFGDPL 2400
Qy 2401 SFGSWTVSEASEDDVCCSMSTYTGTALITPCAAEESKLPINPLNSLRLHNNMVYAT 2460
Db 2401 SFGSWTVSEAGEDVCCSMSTYTGTALITPCAAEESKLPINPLNSLRLHNSMVYST 2460
Qy 2461 SRSASLRQKKVTFDRLOVLDHRYDVLKEMKAKASTVAKLLSIEACKLTTPHSAKSP 2520
Db 2461 SRSASLRQKKVTFDRLOVLDHRYDVLKEMKAKASTVAKLLSIEACKLTTPHSAKSP 2520
Qy 2521 GYGAKDVRLSSRAVNHRSVWEDLLEDTETIDTITMAKSVFVQPEKGRKEARLIV 2580
Db 2521 GYGAKDVRLSSRAVNHRSVWEDLLEDTETIDTITMAKSVFVQPEKGRKEARLIV 2580
Qy 2581 FPDLVRCVKMALVDVYSTLPQAVNGSSYGFQYSPKQRFVFLVNTWKSCKCPMGFSYDT 2640
Db 2581 FPDLVRCVKMALVDVYSTLPQAVNGSPYGFQYSPQRFVFLVNTWKSCKCPMGFSYDT 2640
Qy 2641 RCFDSTVTSERLVBESYQCDLAPARQAIRSLTERLYIGGPLTNSKGQCGVRRORA 2700
Db 2641 RCFDSTVENDIRVBESYQCDLAPARQAIRSLTERLYIGGPLTNSKGQCGVRRORA 2700
Qy 2701 SGVLTSCGNTLTCYLKATAACRAAKLQDCVWLVNGDDLVICISAGTQEDAAALRASTE 2760
Db 2701 SGVLTSCGNTLTCYLKATAACRAAKLQDCVWLVNGDDLVICISAGTQEDAAALRASTE 2760
Qy 2761 AMTRYSAAPGDPPEYDELEITSCSNVSVAHDAAGKRVYLTDRDPTPLARAWEIAR 2820
Db 2761 AMTRYSAAPGDPPEYDELEITSCSNVSVAHDAAGKRVYLTDRDPTPLARAWEIAR 2820
Qy 2821 HTPVNSWLGNIIMVAPTUWARMILMTHFFSILLAQEOLKALDCQIYACYSIEPLDLQP 2880
Db 2821 HTPVNSWLGNIIMVAPTUWARMILMTHFFSILLAQEOLKALDCQIYACYSIEPLDLQP 2880
Qy 2881 IIERLHGLSAFTLSHSPGEINRVASCLKLGVPPLRWHRARSVRAKLISQGRAATC 2940
Db 2881 IIERLHGLSARLSHSPGEINRVASCLKLGVPPLRWHRARSVRAKLISQGRAATC 2940
Qy 2941 GRVLENWARTKLTIPAPASQDLGWFVAGYSGSDIYHSLGRAPRPWFPLCILLISV 3000
Db 2941 GKYLEFWAVTKLKITPAPASQDLGWFVAGYNGGDIYHSLGRAPRPWFPLCILLISV 3000
Qy 3001 CVGIYLLPNR 3010
Db 3001 CVGIYLLPNR 3010

RESULT 2
GNVWTC
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: A38465
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991
A:Title: Structure and organization of the hepatitis C virus genome isolated from human A:Reference number: A38465; MUID:91140698; PMID:1847440
A:Accession: A38465
A:Molecule type: genomic RNA
A:Residues: 1-3010 <TAX>
A:Cross-references: EMBL:MS8335; NID:g329770; PIDN:AAA72945.1; PID:g323771
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
```

F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEAH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 95.3%; Score 15253; DB 1; Length 3010;
Best Local Similarity 94.4%; Pred. No. 0;
Matches 2840; Conservative 78; Mismatches 92; Indels 0; Gaps 0;

```
Qy 1 MSTNPKQKRTKNTNRRPQVQKTFGGQIVGGVYLLPRRPRIGVRAATRKASRSQPRG 60
Db 1 MSTNPKQKRTKNTNRRPQVQKTFGGQIVGGVYLLPRRPRIGVRAATRKASRSQPRG 60
Qy 61 RRQIPKARRPGRGAWAQPGVPMPLYGNEGLGWAGWLLSPRGSPSWGPTDPRRSNLLG 120
Db 61 RRQIPKARRPGRGAWAQPGVPMPLYGNEGLGWAGWLLSPRGSPSWGPTDPRRSNLLG 120
Qy 121 KVIDTLTCGFADLMGYIPLVGAPLGGARALAHGVRLVLEGVNATGNLPCCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGARALAHGVRLVLEGVNATGNLPCCSFSIFLLA 180
Qy 181 LLSCLTTPASAYVNRVNSGIYHVNTDCNSISIVVEAADVIMHTEGCVPCVQEGNSSRCW 240
Db 181 LLSCLTTPASAYVNRVNSGIYHVNTDCNSISIVVEAADVIMHTEGCVPCVQEGNSSRCW 240
Qy 241 ALPTLAAARNASVPTTIRRHVDLLVCTAFAFCSAMVYVGLDCCGSLFLYSOLFSPRRHET 300
Db 241 ALPTLAAARNVITPTTIRRHVDLLVGAAPCSAMVYVGLDCCGSLFLYSOLFSPRRHET 300
Qy 301 VQDCNCSIIYGHVSGHRMAMMMNWSPTTALVVSQLLRIPQAVDMVAGAHGVLGAGLA 360
Db 301 LQDCNCSIIYGHVSGHRMAMMMNWSPTTALVVSQLLRIPQAVDMVAGAHGVLGAGLA 360
Qy 361 YYSVNGWAKVILVALLFAGVDGTHHTGRVAGHTTSGTSLFSSGASQKIQLNTNGSW 420
Db 361 YYSVNGWAKVILVALLFAGVDGTHHTGRVAGHTTSGTSLFSSGASQKIQLNTNGSW 420
Qy 421 HINRTALNCDSIQTGFFAALFYAHKPNSSCGPERVASCRIPIWFAQCGMPIYTKENSS 480
Db 421 HINRTALNCDSIQTGFFAALFYTHFNSSGCGPERVACQRTIDKFDQGWGPIYABSSRS 480
Qy 481 DQRPYCHYAPREPCGVVFPASQVGPVYCFTPSPVAVVGTTRDSGVYTVSGENETVMLLN 540
Db 481 DQRPYCHYAPREPCGVVFPASQVGPVYCFTPSPVAVVGTTRDSGVYTVSGENETVMLLN 540
Qy 541 NTRPPQGNWPGCTWMSSTGTCTCGGPPCNIGGVGNRTLICPTDCFRKPIEATYTKCGSG 600
Db 541 NTRPPQGNWPGCTWMSSTGTCTCGGPPCNIGGVGNRTLICPTDCFRKPIEATYTKCGSG 600
Qy 601 PMLTPLCLVDYVRLWHYPTCLNFSIFKVMVYVGVVEHRLNAACNMTGRONLEDRDS 660
Db 601 PMLTPLCMVDYVRLWHYPTCLNFTIFKVMVYVGVVEHRLNAACNMTGRONLEDRDP 660
Qy 661 ELSPLLSTTETWQILPCAPFTLTPALSTGLHQLHNTVDVQYLYGVGSFAVFAIKWEYIL 720
Db 661 ELSPLLSTTETWQVLPSTLTPALSTGLHQLHNTVDVQYLYGVGSFAVFAIKWEYVL 720
Qy 721 LLELLADARVCAKLMMLLIAQAEALNVLNLAASVAGAHGILSLVLPVFFCAAWYIKG 780
Db 721 LLELLADARVCAKLMMLLIAQAEALNVLNLAASVAGAHGILSLVLPVFFCAAWYIKG 780
Qy 781 RIAPGAAYAFYGVWPLLLLLALLPPAYALDRMAASCGGAVLVGLVFTLSPYKVFET 840
Db 781 RLVPGTATYALYGVWPLLLLLALLPPAYALDRMAASCGGAVLVGLVFTLSPYKVFET 840
Qy 841 RLIVWLQYITRAEAHMQVWVPPVPLNVRGGRDAIILITCAVHPBELIDITKLLAILGLPLM 900
```

Db 841 RL1WMLQVFTTRAEADLHVPIPLNARCGRDAILLIMCAVHPBLIFDITKILILAILGLPM 900
QY 901 VLQAGITRVFVRAQGLIRACMLVRKVAGHYQVMFMKLGALTGTYYNHLPLRDWA 960
Db 901 VLQAGITRVFVRAQGLIRACMLVRKVAGHYQVMFMKLGALTGTYYNHLPLRDWP 960
QY 961 HAGLRDLAVAPVVPVFSAMETKVTIWGADTAACGDIILGLPVSARRRKEIFLGPADSLEG 1020
Db 961 RAGLRDLAVAPVVPVFSAMETKVTIWGADTAACGDIILGLPVSARRRKEIFLGPADSLEG 1020
QY 1021 QGWILLAPITAYSQOTRGVLGCLITSLTGRDNQVGEVQVSTATQSFLATCINGVCWT 1080
Db 1021 RGLRLLAPITAYSQOTRGVLGCLITSLTGRDNQVGEVQVSTATQSFLATCINGVCWT 1080
QY 1081 VYHGAGSXTLAGPKPTOMXTNVLDLGVHQAAPGARSMPTPCGSGSDLYLVTRHADVI 1140
Db 1081 VYHGAGSXTLAGPKPTOMXTNVLDLGVHQAAPGARSMPTPCGSGSDLYLVTRHADVI 1140
QY 1141 PVRRGDSRGSLSPRPVSVYKGSGGPCLLCPSGHVGVPPAAVCTRGVAKAVDPFIPVES 1200
Db 1141 PVRRGDSRGSLSPRPVSVYKGSGGPCLLCPFGHAVGIFPAAVCTRGVAKAVDPFIPVES 1200
QY 1201 METTMRSPVFTDNSTPPAVPOTFOVAHLHAPTGSKSTKVPAAVAAQYKVLVNPVAA 1260
Db 1201 METTMRSPVFTDNSTPPAVPOTFOVAHLHAPTGSKSTKVPAAVAAQYKVLVNPVAA 1260
QY 1261 TLGFGAYMSKAHGIDPNIRTVRTITGSGSTYTYGKFLADGCGSGAYDIIICDCHS 1320
Db 1261 TLGFGAYMSKAHGIDPNIRTVRTITGAPVTSYTYGKFLADGCGSGAYDIIICDCHS 1320
QY 1321 TDSSTILGCTVLQDAETAGARLVLATATPPGSVTVZPHNIEIGLSNNGEIPFYGKAI 1380
Db 1321 TDSSTILGCTVLQDAETAGARLVLATATPPGSVTVZPHNIEIEVALSNTGEIPFYGKAI 1380
QY 1381 PIERAIGRHLI PCHSKKKDBELAAKTGLGLNAVAYRGLDVSVIPIGDVVVVAATDAL 1440
Db 1381 PIERAIGRHLI PCHSKKKDBELAAKLSGLINAVAYRGLDVSVIPIGDVVVVAATDAL 1440
QY 1441 MTGFTGFDSDVNDCTVOTVDFSLDPTFTTETTTVPQAVSSORRGTRGSGSIYR 1500
Db 1441 MTGFTGFDSDVNDCTVOTVDFSLDPTFTTETTTVPQAVSSORRGTRGSGSIYR 1500
QY 1501 FVTPGERPSGMFDSVLCCEYDAGCAWYELTFAETSVRLRAYLNTPLPVQCQDHLFEWES 1560
Db 1501 FVTPGERPSGMFDSVLCCEYDAGCAWYELTFAETSVRLRAYLNTPLPVQCQDHLFEWES 1560
QY 1561 VFTGLTHIDAHFLSCTQAGDNPPYLVAYQATVCARAQAPPPSDQWQKCLIRLKPFLHG 1620
Db 1561 VFTGLTHIDAHFLSCTQAGDNPPYLVAYQATVCARAQAPPPSDQWQKCLIRLKPFLHG 1620
QY 1621 PTLPLYLGAQVQNEVILTHPIITKYIMACMGADLEVVTSTWLVGVVLAALAAAYCLITGSV 1680
Db 1621 PTLPLYLGAQVQNEVILTHPIITKYIMACMGADLEVVTSTWLVGVVLAALAAAYCLITGSV 1680
QY 1681 VIIVGRITLSKPAVVPDREVLVYQFDEMEBCASOLPYIEQOMLAEQFKOKA-JGLLOTAT 1740
Db 1681 VIIVGRITLSKPAVVPDREVLVYQFDEMEBCASOLPYIEQOMLAEQFKOKA-JGLLOTAT 1740
QY 1741 KQAEAAAPVVEKWRALETTWAKEMWNFIISGIVLAGISTLPGNPAIASIMATASITSP 1800
Db 1741 KQAEAAAPVVEKWRALETTWAKEMWNFIISGIVLAGISTLPGNPAIASIMATASITSP 1800
QY 1801 LTTQNTLLFNILGWSVAAQALAPPSAASAFVAGIAGAGVAGSIGLGVKLVILLAGYGAGVA 1860
Db 1801 LTTQNTLLFNILGWSVAAQALAPPSAASAFVAGIAGAGVAGSIGLGVKLVILLAGYGAGVA 1860
QY 1861 GALVAFKVMGSEVPSTEDLNNLIPAILSPGALVGVVVCVCAAILRRHVGPBGAGVQVMNRLI 1920
Db 1861 GALVAFKVMGSEVPSTEDLNNLIPAILSPGALVGVVVCVCAAILRRHVGPBGAGVQVMNRLI 1920
QY 1921 AFASRGNHVSPTHVVPESDAAAVTQTLSSITQTLKRLHQMINEDCSTPCSGSMLRDV 1980
Db 1921 AFASRGNHVSPTHVVPESDAAAVTQTLSSITQTLKRLHQMINEDCSTPCSGSMLRDV 1980

QY 1981 WDMICTVLIDFKTWMQSKLLPRLPGVPFLSCQRYGKVMRGDGMOTPCGQAQIAGHVK 2040
Db 1981 WDMICTVLIDFKTWMQSKLLPRLPGVPFLSCQRYGKVMRGDGMOTPCGQAQIAGHVK 2040
QY 2041 NGSMRIVGPETCNTWHTGTFPINAYTTGCTPSPAPNYSBALWRAAEZVVEVTRVGDFH 2100
Db 2041 NGSMRIVGPETCNTWHTGTFPINAYTTGCTPSPAPNYSBALWRAAEZVVEVTRVGDFH 2100
QY 2101 YVTGMTTNDVKCQCPAPEFFTEVDGVLHRYAPACKPILLREDVTFQVGLNLYLVGSOL 2160
Db 2101 YVTGMTTNDVKCQCPAPEFFTEVDGVLHRYAPACKPILLREDVTFQVGLNLYLVGSOL 2160
QY 2161 PCBPEDDVTVLTMLTDPHSHTATETAKRLARSPPSLASSASQUSAPSLKATCTTHHD 2220
Db 2161 PCBPEDDVTVLTMLTDPHSHTATETAKRLARSPPSLASSASQUSAPSLKATCTTHHD 2220
QY 2221 SPADLIIEANLWROBMGNITRESEBENKVVILDSFEPHAEGDDEIEISVAABILEKSRK 2280
Db 2221 SPADLIIEANLWROBMGNITRESEBENKVVILDSFEPHAEGDDEIEISVAABILEKSRK 2280
QY 2281 FPSALPIWARPNDYNPPLLESWKDPDYPVPVWHGCPPLPTKAPPVPPRRKRTVVLTESNV 2340
Db 2281 FPSALPIWARPNDYNPPLLESWKDPDYPVPVWHGCPPLPTKAPPVPPRRKRTVVLTESNV 2340
QY 2341 SSALABLATKTFSSGSSSAVDSGTATALPDLASDDGKGSDBESYSSMPPLEGEPPDPL 2400
Db 2341 SSALABLATKTFSSGSSSAVDSGTATALPDQASDDGKGSDBESYSSMPPLEGEPPDPL 2400
QY 2401 SDGSWSTVSEASEDDVVCSSMSYTWTCALITPCAAEESKLPINPLNSLLRHHNMVYATT 2460
Db 2401 SDGSWSTVSEASEDDVVCSSMSYTWTCALITPCAAEESKLPINPLNSLLRHHNMVYATT 2460
QY 2461 SRGASLRQKXVTPDRLOVLDHYRUVLKNKAKASTVAKALLSIEACKLTPPHSAKSXF 2520
Db 2461 SRGASLRQKXVTPDRLOVLDHYRUVLKNKAKASTVAKALLSIEACKLTPPHSAKSXF 2520
QY 2521 GYKAKVRNLSSEAVNHIHSEVEDLLEDTEPIDTITMAKSEVFCVQPEKGGKPARLIV 2580
Db 2521 GYKAKVRNLSSEAVNHIHSEVEDLLEDTEPIDTITMAKSEVFCVQPEKGGKPARLIV 2580
QY 2581 FPDPLGVRCCKMALYDWSLTPQAVMGSSYGFQYSPKQVFEFLVNTWKSKCPMGFSYDT 2640
Db 2581 FPDPLGVRCCKMALYDWSLTPQAVMGSSYGFQYSPKQVFEFLVNTWKSKCPMGFSYDT 2640
QY 2641 RCFDSTVTSDSIRVESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQCGYRCRA 2700
Db 2641 RCFDSTVTSDSIRVESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQCGYRCRA 2700
QY 2701 SGVLTTSCGNTLTCYLKATAACRAAKLODCTMLVNGDDLVWICESAGTOEDAAALRAFT 2760
Db 2701 SGVLTTSCGNTLTCYLKATAACRAAKLODCTMLVNGDDLVWICESAGTOEDAAALRAFT 2760
QY 2761 AMTRYSAPEGDPQPEYDLELITSCSSNVSVADHASKRVTYLTDRPTTPLARAWEIAR 2820
Db 2761 AMTRYSAPEGDPQPEYDLELITSCSSNVSVADHASKRVTYLTDRPTTPLARAWEIAR 2820
QY 2821 HTPINSWLNILMYAPTLWARMLMTHPFSIILAOBKLKALDCCQIYGACYSIEPLDLPQ 2880
Db 2821 HTPINSWLNILMYAPTLWARMLMTHPFSIILAOBKLKALDCCQIYGACYSIEPLDLPQ 2880
QY 2881 IIERLHGLSAFTLHYSYSGEINRVASCLRKLVPPILRTWRHARSVRACKLSQGGRAATC 2940
Db 2881 IIERLHGLSAFTLHYSYSGEINRVASCLRKLVPPILRTWRHARSVRACKLSQGGRAATC 2940
QY 2941 GRYLFWNAVTKLXLTPIPAASQDLDSGMFVAGYSGGDIYHLSLRARPEWFLCLLLSV 3000
Db 2941 GRYLFWNAVTKLXLTPIPAASQDLDSGMFVAGYSGGDIYHLSLRARPEWFLCLLLSV 3000
QY 3001 GVGIYLLPNR 3010
Db 3001 GVGIYLLPNR 3010

RESULT 3
A45573
Genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome protein NS4; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
A:Reference number: A45573; M01D:92295714; PMID:1318627
A:Accession: A45573
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>
A:Cross-references: GB:D11168; GB:D01171; MID:g221612; PIDN:BA001943.1; P-ID:g221613
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin
F:2-115/Product: capsid protein C #status: predicted <CPC>
F:116-191/Product: envelope protein M #status: predicted <BPM>
F:192-389/Product: major envelope protein E #status: predicted <ME3>
F:390-728/Product: nonstructural protein NS1 #status: predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status: predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome protein NS3 #status: predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1662/Product: nonstructural protein NS4a #status: predicted <N4a>
F:1863-2013/Product: nonstructural protein NS4b #status: predicted <N4b>
F:2014-3010/Product: nonstructural protein NS5 #status: predicted <NS5>

Query Match 95.2%; Score 15236; DB 1; Length 3010;
Best Local Similarity 94.1%; Pred. No. 0;
Matches 2833; Conservative 83; Mismatches 94; Indels 0; Gaps 0;

QY 1 MSTNPKPORKTRNNRRPDQVKKFGGQIVGGVYLLPRRGRLGVRAIRKASERSQRRG 60
DB 1 MSTNPKPORKTRNNRRPDQVKKFGGQIVGGVYLLPRRGRLGVRAIRKASERSQRRG 60

QY 61 RROPIKARRBEGRAWAQGYVWPIYGNELGWAGWLLSPGRSPSWGPTDPRRSRNLG 120
DB 61 RROPIKARRBEGRAWAQGYVWPIYGNELGWAGWLLSPGRSPSWGPTDPRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVYVLEDGNYATGNLPGCSPIFLIA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVYVLEDGNYATGNLPGCSPIFLIA 180

QY 181 LLSCLTIPASAYEVNRVNSGIYHVTNDCNSSTIYVEAADVIMHTPCVPCVORGRSSRCWV 240
DB 181 LLSCLTIPASAYEVNRVNSGIYHVTNDCNSSTIYVEAADVIMHTPCVPCVORGRSSRCWV 240

QY 241 ALTPTLAARNASVPTTTRRHVDLLVGTAAFCSAMVYGDLCGSIPLVSQLTFSPRRHET 300
DB 241 ALTPTLAARNASVPTTTRRHVDLLVGTAAFCSAMVYGDLCGSIPLVSQLTFSPRRHET 300

QY 301 VQDCNCSITPGHVSHERMAWDMNWSPTTALVYSQILRIIQAADVYAGAHGWVLAGLA 360
DB 301 VQDCNCSITPGHVSHERMAWDMNWSPTTALVYSQILRIIQAADVYAGAHGWVLAGLA 360

QY 361 YYSMVGNWAKVLIIVALLFAGVDGETHTTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420
DB 361 YYSMVGNWAKVLIIVALLFAGVDGETHTTGRVAGHTTSGTSLFSSGASQKIQLVNTNGSW 420

QY 421 HINTALNCNLSQGTFFAALFYAHKFNSSGCPERMASCRPIDMFAQWGPIITTKPNSS 480
DB 421 HINTALNCNLSQGTFFAALFYAHKFNSSGCPERMASCRPIDMFAQWGPIITTKPNSS 480

QY 481 DORPYCNHAYPRCGVGPASQVCGPVYCFTPSPVVTGTTDRSGAFTYNWGANETDVLLLN 540

DB 481 DORPYCNHAYPRCGVGPASQVCGPVYCFTPSPVVTGTTDRSGAFTYNWGANETDVLLLN 540

QY 541 NTRPPQGNWFGCTWNNSTGFTKTCGGPPCNIGGVGNRTLIPTDPCRKHPEATYTKCGSG 600
DB 541 NTRPPQGNWFGCTWNNSTGFTKTCGGPPCNIGGVGNRTLIPTDPCRKHPEATYTKCGSG 600

QY 601 PWLTPRCVLVDYPRYLWHPYPCITNFSGIFKVMYVGVGVEHRLNAACNWTSGECNLEDRRS 660
DB 601 PWLTPRCVLVDYPRYLWHPYPCITNFSGIFKVMYVGVGVEHRLNAACNWTSGECNLEDRRS 660

QY 661 ELSPILLSTTEWQTLPCSFITLPAALSTGLIHLHQNIVDVQYLYGVGSFAVFAIKWEYIL 720
DB 661 ELSPILLSTTEWQTLPCSFITLPAALSTGLIHLHQNIVDVQYLYGVGSFAVFAIKWEYIL 720

QY 721 LLFLLADARVACACLWMMILLITAOBAALLENVLNNAASVAGAHGILSLVLFCAAWYIKG 780
DB 721 LLFLLADARVACACLWMMILLITAOBAALLENVLNNAASVAGAHGILSLVLFCAAWYIKG 780

QY 781 RLAPGAAYAFYGVMPPLLLLLLPPRAYALDREMAASCGGAVLVGLVFLTISPYKVELT 840
DB 781 RLAPGAAYAFYGVMPPLLLLLLPPRAYALDREMAASCGGAVLVGLVFLTISPYKVELT 840

QY 841 RLIVPGAAVALYGVWPLLLLLLPPRAYAMOREMAASCGGVFVGLILLITLSPHYKVFIA 840
DB 841 RLIVPGAAVALYGVWPLLLLLLPPRAYAMOREMAASCGGVFVGLILLITLSPHYKVFIA 840

QY 841 RLIIWMLQVITRAEAHLCVWVPPNLNVRGGRDAIILLTCAVHPELIFDTIKLLAILGLPM 900
DB 841 RLIIWMLQVITRAEAHLCVWVPPNLNVRGGRDAIILLTCAVHPELIFDTIKLLAILGLPM 900

QY 901 VLOAGITVPPVFAOGLIRACMLVRKVACGHYQVMFKLGAITGTIVYVNHLPRLDWA 960
DB 901 VLOAGITVPPVFAOGLIRACMLVRKVACGHYQVMFKLGAITGTIVYVNHLPRLDWA 960

QY 961 HAGLRDLAVAVPVVFSAMETKVITWGAADTAACGDIILGLPVSABERGKEIFLGADSLG 1020
DB 961 HAGLRDLAVAVPVVFSAMETKVITWGAADTAACGDIILGLPVSABERGKEIFLGADSLG 1020

QY 1021 QGWLLAPITAYSOOTRGVLCIIITSLTRDKNVEGEVQVVSATQSLATCINGVCWT 1080
DB 1021 QGWLLAPITAYSOOTRGVLCIIITSLTRDKNVEGEVQVVSATQSLATCINGVCWT 1080

QY 1081 VYHGAGSKTLAGPKGPIQMTYNTVDLIVGWAOPPGARSMTPCSGSSDLVLVTRHADVI 1140
DB 1081 VYHGAGSKTLAGPKGPIQMTYNTVDLIVGWAOPPGARSMTPCSGSSDLVLVTRHADVI 1140

QY 1141 PVRRGSGRGLLSRPRVSYLKGSGGFLPCSPGVVGPRAAVCTRGVAKAVDFIPVES 1200
DB 1141 PVRRGSGRGLLSRPRVSYLKGSGGFLPCSPGVVGPRAAVCTRGVAKAVDFIPVES 1200

QY 1201 METTMRSPVFDNSPPAVPOTFQVAHLHAPTSGSKSTKVPAAVAAQGYKVLVNPVAA 1260
DB 1201 METTMRSPVFDNSPPAVPOTFQVAHLHAPTSGSKSTKVPAAVAAQGYKVLVNPVAA 1260

QY 1261 TLGRGAYMSKAHGDPMIRTCVRIITGGSTIYTYGKFLADGCGSGAYDIIICDECHS 1320
DB 1261 TLGRGAYMSKAHGDPMIRTCVRIITGGSTIYTYGKFLADGCGSGAYDIIICDECHS 1320

QY 1321 TDSITLIGITGLDQAEATAGARLVILATATPPGSVTVPHNIBERIGLSNNGEIPFYGKAI 1380
DB 1321 TDSITLIGITGLDQAEATAGARLVILATATPPGSVTVPHNIBERIGLSNNGEIPFYGKAI 1380

QY 1381 PLEAKGGRHLIIFCHSKKKDELAALKLGLGINAVAYRGLDVSIVPTSGLVIVATDAL 1440
DB 1381 PLEAKGGRHLIIFCHSKKKDELAALKLGLGINAVAYRGLDVSIVPTSGLVIVATDAL 1440

QY 1441 MTGTGDFDSVDCNTCVTCTVDFSLDPTETIETTVTPQDAVRSQRGHTGRGSIYR 1500
DB 1441 MTGTGDFDSVDCNTCVTCTVDFSLDPTETIETTVTPQDAVRSQRGHTGRGSIYR 1500

QY 1501 FVTPGERPSGWFSSVLCECYDAGCAWVELTPAETSVLRAYLNTPLPCVCDHLEPWS 1560
DB 1501 FVTPGERPSGWFSSVLCECYDAGCAWVELTPAETSVLRAYLNTPLPCVCDHLEPWS 1560

QY 1561 VFTGLTHIDAEFTSQTKAGDNFPYLVAQATVCAQAQPPPSQDMWKCLIKLPTLHG 1620
DB 1561 VFTGLTHIDAEFTSQTKAGDNFPYLVAQATVCAQAQPPPSQDMWKCLIKLPTLHG 1620


```
QY 1621 PTPLLYRLGAVQNEVILTHPTIKYIMAGMADLEVVSTWVLGGVLAALAAAYCLTTGSV 1680
Db 1621 PTPLLYRLGAVQNEVILTHPTIKYIMAGMADLEVVSTWVLGGVLAALAAAYCLTTGSV 1680
QY 1681 VIVGRIILSKPAPVPPREVLYQFEDMEECASOLPYIEQVMQALAEQFKKALGLLQAT 1740
Db 1681 VIVGRIILSKPAPVPPREVLYQFEDMEECASOLPYIEQVMQALAEQFKKALGLLQAT 1740
QY 1741 KQAEAAAPVVEKWEALETFWAKHWNFIISGIVLAGISYLPGNPAIASLMAFTASITSP 1800
Db 1741 KQAEAAAPVVEKWEALETFWAKHWNFIISGIVLAGISYLPGNPAIASLMAFTASITSP 1800
QY 1801 LTTQNTLLENILGWAQAQAPSAASAFVAGAGIAGAAGVSGIGLGVLYVDILAGYGAGVA 1860
Db 1801 LTTQNTLLENILGWAQAQAPSAASAFVAGAGIAGAAGVSGIGLGVLYVDILAGYGAGVA 1860
QY 1861 GALVAPKVMGSEVSTEDLVNLLPAILSPGALVGVVCAALRRHVGPFGAGVQWNRLLI 1920
Db 1861 GALVAPKVMGSEVSTEDLVNLLPAILSPGALVGVVCAALRRHVGPFGAGVQWNRLLI 1920
QY 1921 AFASRGNVSPTHVVPESDAAARVTQILSLTITQLLKLHOWNEDCSTPCSGSMLRDV 1980
Db 1921 AFASRGNVSPTHVVPESDAAARVTQILSLTITQLLKLHOWNEDCSTPCSGSMLRDV 1980
QY 1981 WDMICTVLTDFKTLWQSKLLPRIPGVFPLSCQRYKGVWRGDMQTMOTTCFQGAQIAGHVK 2040
Db 1981 WDMICTVLTDFKTLWQSKLLPKLPGVFPFSCQRYKGVWRGDMQTMOTTCFQGAQIAGHVK 2040
QY 2041 NGSNRIIVGPTSCNTWGTGTEPINAATGCTPSPAPNYSALMRVAABEYVEVTRVGDHF 2100
Db 2041 NGSNRIIVGPTSCNTWGTGTEPINAATGCTPSPAPNYSALMRVAABEYVEVTRVGDHF 2100
QY 2101 YVFGMTTNDVKPCQVPAPEFFFEVDGVRHRYAPACKPLLRDVTQVQGLNOVLVGSOL 2160
Db 2101 YVFGMTTNDVKPCQVPAPEFFFEVDGVRHRYAPACKPLLRDVTQVQGLNOVLVGSOL 2160
QY 2161 PCSPEDVTVLTMIDPSSHITAETAKRRLARGSPPSLASSASQLSAPSLKATCTTHD 2220
Db 2161 PCSPEDVTVLTMIDPSSHITAETAKRRLARGSPPSLASSASQLSAPSLKATCTTHD 2220
QY 2221 SPADLTLEALLWFOEMGNITVSESNKVILDSPEPLHAEGRERISVAAILKRSK 2280
Db 2221 SPADLTLEALLWFOEMGNITVSESNKVILDSPEPLHAEGRERISVAAILKRSK 2280
QY 2281 FPSALPIWARPDYNPPLLESKDPDYVPPVYVHGCPLEPTKAPIPPPRRKRTVVLTESNV 2340
Db 2281 FPSALPIWARPDYNPPLLESKSPDYVPPVAVHGCPLPTTGPPIPPPRKRTVVLTESIV 2340
QY 2341 SSALABELATKTFGSSGSAVDSTATLALDPLASDQDGKSDVBSYSMPPLEGEPDPL 2400
Db 2341 SSALABELATKTFGSSGSAVDSTATAPPDQTSDDGKSDVBSYSMPPLEGEPDPL 2400
QY 2401 SDGSWSTVSEAGEDVVCSSMSYTWGALLTPCAARESKLPINPLNSLIRHNNVYATT 2460
Db 2401 SDGSWSTVSEAGEDVVCSSMSYTWGALLTPCAARESKLPINPLNSLIRHNNVYATT 2460
QY 2461 SRASLRQKVVTDRLQVLDHHDYDLVKEMKAKASTVKAKILSIEACKLTPPHSAKSKF 2520
Db 2461 SRASLRQKVVTDRLQVLDHHDYDLVKEMKAKASTVKAKILSIEACKLTPPHSAKSKF 2520
QY 2521 GYKAKOVNLSRAVNHIRSVWEDLLEDTEPTDITIMAKSEVFCVQPEKGRKPARLIV 2580
Db 2521 GYKAKOVNLSRAVNHIRSVWEDLLEDTEPTDITIMAKSEVFCVQPEKGRKPARLIV 2580
QY 2581 FPDGLGVACEKMAIYDVSTLPAVMSSYGFQYSPKQVFEFLVNTWKSCKCWGSGSYDT 2640
Db 2581 FPDGLGVACEKMAIYDVSTLPAVMSSYGFQYSPKQVFEFLVNTWKSCKCWGSGSYDT 2640
QY 2641 RCTDSTVTESDIRVESIYOCCLDAPARQAIKSLTERLYIGGFLINSKQNGCYRCRA 2700
Db 2641 RCTDSTVTESDIRVESIYOCCLDAPARQAIKSLTERLYIGGFLINSKQNGCYRCRA 2700
```

```
QY 2701 SGVLTTCGNTLCYLKATAACRAAKLQDCTMLVNGDDLWICESAGTODAAALAFTE 2760
Db 2701 SGVLTTCGNTLCYLKATAACRAAKLQDCTMLVNGDDLWICESAGTODAAALAFTE 2760
QY 2761 AMTRYSAAPPDPPQPEYDLELITSCSSNVSVAHDAAGKAVVYLTRDPTTFLAAAAWETAR 2820
Db 2761 AMTRYSAAPPDPPQPEYDLELITSCSSNVSVAHDAAGKAVVYLTRDPTTFLAAAAWETAR 2820
QY 2821 HTPINSWNLGNIIMYAPTLWARMILMTHFFSILLAQOLEKALDCQIYGACYSIEPLDLQ 2880
Db 2821 HTPINSWNLGNIIMYAPTLWARMILMTHFFSILLAQOLEKALDCQIYGACYSIEPLDLQ 2880
QY 2881 IIRERHGLSAFTLHYSYSGEINRVASCLKLGVPPLRTWRHRARSVRKILSOGGAAATC 2940
Db 2881 IIRERHGLSAFTLHYSYSGEINRVASCLKLGVPPLRTWRHRARSVRKILSOGGAAATC 2940
QY 2941 GRYLFNVAWRTKLTPIPAASQDLSGMFVAGYGGDIYHLSLRARPWFPLCLLLSV 3000
Db 2941 GRYLFNVAWRTKLTPIPAASQDLSGMFVAGYGGDIYHLSLRARPWFPLCLLLSV 3000
QY 3001 GVGIVYLLPNR 3010
Db 3001 GVGIVYLLPNR 3010

RESULT 4
GNWTV
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A:Reference number: A40244; MUID:92230206; PMID:131449
A:Accession: A40244
A:Molecule type: genomic RNA
A:Residues: 1-3010 <CHE>
A:Cross-references: GB:M84754
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
P:1-115/Product: capsid protein C #status predicted <CP>
P:116-191/Product: major envelope protein M #status predicted <EM>
P:192-389/Product: major envelope protein E #status predicted <EP>
P:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
P:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
P:1007-1615/Product: hepatitis C virus genome: sequence determination and mapping the
P:1230-1237/Region: nucleotide-binding motif A (P-loop)
P:1312-1317/Region: nucleotide-binding motif A (P-loop)
P:1316-1319/Region: DEXH motif
P:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
P:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
P:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
P:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207
```

Query Match 94.1%; Score 15062; DB 1; Length 3010;

Best Local Similarity 93.3%; Pred. No. 0;

Matches 2807; Conservative 81; Mismatches 122; Indels 0; Gaps 0;

QY 1 MSTNPKPQKTKNTNRRPDQVKFPGGQIVGGVLLPRRGPRLGVATRKASERSQPRG 60

Db 1 MSTNPKPQKTKNTNRRPDQVKFPGGQIVGGVLLPRRGPRLGVATRKATWERSQPRG 60

QY 61 RQQIPKARPPGRWAQYQYWPVLYGNEGLWAGWLSPROSRFSWQPTDPRRSRLG 120

Db 61 RQQIPKARPPGRWAQYQYWPVLYGNEGLWAGWLSPROSRFSWQPTDPRRSRLG 120

QY 121 KVIDLTCGFADLMGYIPLVGAFLGVARALAHGVRLSDGVNYATGNLPGCSFIFLLA 180

Db 121 KVIDLTCGFADLMGYIPLVGAFLGVARALAHGVRLSDGVNYATGNLPGCSFIFLLA 180

181 LLSCLTIPASAYEVNVSIGIYHVNDSCNSIYVEAADVIMHTPCVPCVQEGNSSRCWV 240
181 LLSCLTIPASAYEVNVSIGIYHVNDSCNSIYVEAADVIMHTPCVPCVQEGNSSRCWV 240
241 ALPTLTAARNASVPTTIRREVDLLVGTAFCSAMYVGDLCGSIFLVSLFTSPREHET 300
241 ALPTLTAARNASVPTTIRREVDLLVGTAFCSAMYVGDLCGSIFLVSLFTSPREHET 300
301 VQDCNCSIIYGHVSIGHRMAMDMXWSPPTALVVSQILIRIPQAVVDVAGAHVGLAGLA 360
301 VQDCNCSIIYGHVSIGHRMAMDMXWSPPTALVVSQILIRIPQAVVDVAGAHVGLAGLA 360
361 YYSVGNWAKVLIYALLFAGVDEGTHTRGVAGHTTSQFTSLFSSGASQXQLINTNGSW 420
361 YYSVGNWAKVLIYALLFAGVDEGTHTRGVAGHTTSQFTSLFSSGASQXQLINTNGSW 420
421 HINRTALNCNDSLOTGFPAALFYAHKNSSGCPPRMASCRPIDWFAQGWGPIYTKFNSS 480
421 HINRTALNCNDSLOTGFPAALFYAHKNSSGCPPRMASCRPIDWFAQGWGPIYTKFNSS 480
481 DQRPYCHYAPRPGVVPASQVGPVCFPTSPVVGTTDRSGVPTYSWGENETDVLIN 540
481 DQRPYCHYAPRPGVVPASQVGPVCFPTSPVVGTTDRSGVPTYSWGENETDVLIN 540
541 NTRPPQGNWFGCTWNNSTGFTKTCGGPPCNIGGVGNRTLICPTDCFRKHPEATYTKCGSG 600
541 NTRPPQGNWFGCTWNNSTGFTKTCGGPPCNIGGVGNRTLICPTDCFRKHPEATYTKCGSG 600
601 PWLTPRCCLVDYPIELWNYPCLTNPSIFKVMYVGVGEHRLNAAWTRGRCNLEDRDRS 660
601 PWLTPRCCLVDYPIELWNYPCLTNPSIFKVMYVGVGEHRLNAAWTRGRCNLEDRDRS 660
661 ELSPLLSTTEWQILPCAFTTLPALSTGLHLHONIVDOVLYGVGSFAVSFAIKWXYIL 720
661 ELSPLLSTTEWQILPCAFTTLPALSTGLHLHONIVDOVLYGVGSFAVSFAIKWXYIL 720
721 LLELLADARVACLWVWMLLIAQAEAALENLVLNAAASVAGAHGILSFLVFFCAAWYIKG 780
721 LLELLADARVACLWVWMLLIAQAEAALENLVLNAAASVAGAHGILSFLVFFCAAWYIKG 780
781 ELAPGAAAYGVWYPIILLIALLPAPRAYALDREMAASCGGAVLVGLVPLTSPYKVFLLT 840
781 ELAPGAAAYGVWYPIILLIALLPAPRAYALDREMAASCGGAVLVGLVPLTSPYKVFLLT 840
841 RLILWLOQYFITRAEAHQVWVPLNVNRCGRDAIILLTCAVHPELIPDITKILLAILGLPM 900
841 RLILWLOQYFITRAEAHQVWVPLNVNRCGRDAIILLTCAVHPELIPDITKILLAILGLPM 900
901 VLOAGITRVPYFVRAQGLIRACMLVRKVAGHYVQVFMKLGALTGYVYVNHLPPLRDWA 960
901 VLOAGITRVPYFVRAQGLIRACMLVRKVAGHYVQVFMKLGALTGYVYVNHLPPLRDWA 960
961 HAGLRDLAVAVEPVVFSAMETKVIWGDATACGDIILGLPVSARRKEIFLGPADSLEG 1020
961 HAGLRDLAVAVEPVVFSAMETKVIWGDATACGDIILGLPVSARRKEIFLGPADSLEG 1020
1021 QGWRLLAPITAYSQOTRGVLGCIITSLTGRDKNOVEGEVOVSTATQSFLATCINGVCWT 1080
1021 QGWRLLAPITAYSQOTRGVLGCIITSLTGRDKNOVEGEVOVSTATQSFLATCINGVCWT 1080
1081 VYHAGSKTLIAGPKGPTIQWYTNVNDLVLGVQAPPGARSMTPCSCGSSDLYLVTRHADVI 1140
1081 VYHAGSKTLIAGPKGPTIQWYTNVNDLVLGVQAPPGARSMTPCSCGSSDLYLVTRHADVI 1140
1141 PVRRGDSRGLLSLSPRVSYLKGSGGPLLCPSHVNVGVFRAAVCTRGVAKAVDFPVES 1200
1141 PVRRGDSRGLLSLSPRVSYLKGSGGPLLCPSHVNVGVFRAAVCTRGVAKAVDFPVES 1200
1201 METTMRSPVFTDNSTPPAVPOTFOVAHLHAPTGSKGSTKVPAAAYAGYKVLVLPNSVAA 1260
1201 METTMRSPVFTDNSTPPAVPOTFOVAHLHAPTGSKGSTKVPAAAYAGYKVLVLPNSVAA 1260

1261 TLGFGAYMSKAHGIIDNIRITGVRTITTTGSGIYTSYTYGKFLADGGCSGGAYDIILCDECHS 1320
1261 TLGFGAYMSKAHGIIDNIRITGVRTITTTGSGIYTSYTYGKFLADGGCSGGAYDIILCDECHS 1320
1321 TDSTTILGIGTVLDQAEATAGARLVLIATATPPGCVTPVPHPNIBEEIGLSNNGEIPFFYKAI 1380
1321 TDSTTILGIGTVLDQAEATAGARLVLIATATPPGCVTPVPHPNIBEEIGLSNNGEIPFFYKAI 1380
1381 PIBAIXGGHRLIFCHSKKKKDELAALKJGLGNAVAYYRGLDVSVPPIPGDVVVVATDAL 1440
1381 PIBAIXGGHRLIFCHSKKKKDELAALKJGLGNAVAYYRGLDVSVPPIPGDVVVVATDAL 1440
1441 MTGFTGDFSDVDCNCTVITQTVDFSLDPTFTIETTTVPQDAVSRSQRRGTSRGGIYR 1500
1441 MTGFTGDFSDVDCNCTVITQTVDFSLDPTFTIETTTVPQDAVSRSQRRGTSRGGIYR 1500
1501 FVTPGSRPQSGMFDSSVLCBCYDAGCAWYELTTPAETSURLRAYLNTGELPVCQDHLBFWS 1560
1501 FVTPGSRPQSGMFDSSVLCBCYDAGCAWYELTTPAETSURLRAYLNTGELPVCQDHLBFWS 1560
1561 VFTGLTHIDAHFLUSQTKQAGDNFPYLVAYQATVCARAQAAPPSPSDQWVKCLIRLXPTLHG 1620
1561 VFTGLTHIDAHFLUSQTKQAGDNFPYLVAYQATVCARAQAAPPSPSDQWVKCLIRLXPTLHG 1620
1621 PTELLYRLGAVQNEVILTHPIKYIMACMSADLEVVTSWVLVGGVLAALAAAYCLTTGVS 1680
1621 PTELLYRLGAVQNEVILTHPIKYIMACMSADLEVVTSWVLVGGVLAALAAAYCLTTGVS 1680
1681 VIIVGRILISGKPAVVPDREVLYQEPFEMBECAQSOLPIIEQGMQLAEQFKQKALGLQAT 1740
1681 VIIVGRILISGKPAVVPDREVLYQEPFEMBECAQSOLPIIEQGMQLAEQFKQKALGLQAT 1740
1741 KOAAAAAPVVEVSKWRALETFWAKHMNFISGIOYLAGLSTLPGNPAIASLMAFTASITSP 1800
1741 KOAAAAAPVVEVSKWRALETFWAKHMNFISGIOYLAGLSTLPGNPAIASLMAFTASITSP 1800
1801 LTTQNTLLFNILGGWAAQALAPESAASAFVAGIAGAAGVSGISGLGKVLVDILAGYAGVA 1860
1801 LTTQNTLLFNILGGWAAQALAPESAASAFVAGIAGAAGVSGISGLGKVLVDILAGYAGVA 1860
1861 GALVAPKWSGEVPSSTEDLWNLPLPAILSPALVGVVCAAILRRHVGPGEVAVQWNRLLI 1920
1861 GALVAPKWSGEVPSSTEDLWNLPLPAILSPALVGVVCAAILRRHVGPGEVAVQWNRLLI 1920
1921 AFASRGNHVSPTHYVSPESDAARVTOILSSLTITOLLXRLHWINEDCSTPCSGSWLRDV 1980
1921 AFASRGNHVSPTHYVSPESDAARVTOILSSLTITOLLXRLHWINEDCSTPCSGSWLRDV 1980
1981 WDWICTVLDTFKTWLQSKLLPRLPGVFFLSQCRGYKGVWRGDMQITPCGQAQIAGHVK 2040
1981 WDWICTVLDTFKTWLQSKLLPRLPGVFFLSQCRGYKGVWRGDMQITPCGQAQIAGHVK 2040
2041 NGSWRIIVGPRTCNTWHTGTPFNAYTGTCTSPAPNSYRALWRAAABEVETRVYGDH 2100
2041 NGSWRIIVGPRTCNTWHTGTPFNAYTGTCTSPAPNSYRALWRAAABEVETRVYGDH 2100
2101 YVGTMTTNDNVKCCQVPAPEBFFTEVDGVRLHRYAPACKPLLRDVTFOVGLNOLYVGSOL 2160
2101 YVGTMTTNDNVKCCQVPAPEBFFTEVDGVRLHRYAPACKPLLRDVTFOVGLNOLYVGSOL 2160
2161 PCSPBPDVTLTSMILTDPSSHITABETAKRRLARSGPPSLASSASQLSAPSLKATCTTHD 2220
2161 PCSPBPDVTLTSMILTDPSSHITABETAKRRLARSGPPSLASSASQLSAPSLKATCTTHD 2220
2221 SPADLIIEANLLWRQWGGNIIRVSEKENVILDSPEPIHAECDDEREISVAABILKRSK 2280
2221 SPADLIIEANLLWRQWGGNIIRVSEKENVILDSPEPIHAECDDEREISVAABILKRSK 2280
2281 FPSALPIWARPDPNPPLESWKDPDVVPVWVHCPLPPTKAPPIPPPRKRTVWLTVESNV 2340
2281 FPSALPIWARPDPNPPLESWKDPDVVPVWVHCPLPPTKAPPIPPPRKRTVWLTVESNV 2340
2341 SSALABELATKTFGSGSSAVDSCTATALPDLASDDGDKGSDVESYSSMPPLEGEGBDPL 2400

2341 SSALAEATKTFGSSSSAAGSGTATAPPQPSDDGDAGSDVSCSSMPLEGEQDPDL 2400
 2401 SDGSWSTVSEASEEDVCCSMSTWTGALLTPCAABESKLPINPLNSLLRHNNMYATT 2460
 2401 SDGSWSTVSEEDGEGVCCSMSTWTGALLTPCAABESKLPINALNSLLRHNNMYATT 2460
 2461 SRASLRQKQKVTDFRLQVLDHRYDVLKEMKAKASTVAKALLSIEBACKLTPPHSAKSKF 2520
 2461 SRASLRQKQKVTDFRLQVLDHRYDVLKEMKAKASTVAKALLSIEBACKLTPPHSAKSKF 2520
 2521 GYGAQVNRNLSRAVNHIRSWEDLLEDTETPTIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
 2521 GYGAQVNRNLSRAVNHIRSWEDLLEDTETPTIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
 2581 FPLGVRVCKMAYLVVSTLPQAVMSSYGFQYSPQORVEFLVNTWKSCKCPMGFSYDT 2640
 2581 FPLGVRVCKMAYLVVSTLPQAVMSSYGFQYSPQORVEFLVNTWKSCKCPMGFSYDT 2640
 2641 RCFDSTVTESDIRVESIYOCCLAPARQAIISRLTERLYIGGPLTNSKQNGCYRRCRA 2700
 2641 RCFDSTVTESDIRVESIYOCCLAPARQAIISRLTERLYIGGPLTNSKQNGCYRRCRA 2700
 2701 SGVLTSSCGNLTLCYLKATACAAKLODCTMLVNGDDLVVICESAGTCEDAAALRAFTE 2760
 2701 SGVLTSSCGNLTLCYLKATACAAKLODCTMLVNGDDLVVICESAGTCEDAAALRAFTE 2760
 2761 AMTRYSAAPPDPPQPSYDLELITSCSNVSVADHAGSKRVYILTRPTPLAANAETAR 2820
 2761 AMTRYSAAPPDPPQPSYDLELITSCSNVSVADHAGSKRVYILTRPTPLAANAETAR 2820
 2821 HTPVNSWLGNIIMYAPTLWARMLIMTFFSILLAOBLEKALDQCYGACYSIEPLDLQ 2880
 2821 HTPVNSWLGNIIMYAPTLWARMLIMTFFSILLAOBLEKALDQCYGACYSIEPLDLQ 2880
 2881 IERLHGLSFTLHSTPSGEINRVASCLRLKGLVPEPLRTWHRARSVRKLLSQGGAATC 2940
 2881 IERLHGLSFTLHSTPSGEINRVASCLRLKGLVPEPLRTWHRARSVRKLLSQGGAATC 2940
 2941 GRYLFNVAWVKLPIPAASOLDLSGWFVAGYSGDLYHSLSRAPRPFLLCLLLSV 3000
 2941 GRYLFNVAWVKLPIPAASOLDLSGWFVAGYSGDLYHSLSRAPRPFLLCLLLSV 3000
 3001 GVGIVYLLPNR 3010
 3001 GVGIVYLLPNR 3010

RESULT 5
 S18030
 Genome polypeptide - hepatitis C virus (isolate JK1)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Variety: isolate JK1
 C:Date: 19-May-2000 #sequence revision 19-May-2000 #text change 23-Mar-2001
 C:Accession: S18030; S33570; A48332; S18029
 R:Honda, M.; Kaneo, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient
 A:Reference number: S18028
 A:Accession: S18030
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <HON>
 A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
 A:Experimental source: isolate JK1 from an individual
 R:Honda, M.; Kaneo, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
 Arch. Virol. 128, 163-169, 1993
 A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
 A:Reference number: A48332; MUID:93119270; PMID:8380322
 A:Accession: S33570
 A:Molecule type: Genomic RNA
 A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>

A:Cross-references: EMBL:X61591
 A:Note: this sequence is inconsistent with the nucleotide translation
 as TTP, and TTC for residue 43 as Pro, TGG for residue 320 as
 A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
 F:115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted
 F:192-389/Product: major envelope protein E #status predicted <ME>
 F:730-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1230-1237/Region: nucleotide-binding motif B
 F:1230-1237/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4b #status predicted <NA4>
 F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196, 209, 234, 250, 305, 417, 423, 448, 532, 540, 556, 576, 623, 645/Binding site: carbohydrate (Asi)

Query Match 94.0%; Score 15052; DB 1; Length 3010;
 Best Local Similarity 93.2%; Pred. No. 0;
 Matches 2804; Conservative 87; Mismatches 119; Indels 0; Gaps 0;

QY 1 MSTNPKPQKTKENTNRPPQDVKFGGQIVGGVLLRRGPRLGVRAKASERSQRG 60
 DB 1 MSTNPKPQKTKENTNRPPQDVKFGGQIVGGVLLRRGPRLGVRAKASERSQRG 60
 QY 61 RSQIPKARRPRGAWAOGYPWPVLYGNEGLWAGWLLSPRGSRSWGTPTPPRRSRNLG 120
 DB 61 RSQIPKARRPRGAWAOGYPWPVLYGNEGLWAGWLLSPYGSRRWGTPTPPRRSRNLG 120
 QY 121 KVIDTLTCGADLMGYIPLVGAAPLGGAAARLAHGVRLVLDGVNATGNLPGCSFIFLLA 180
 DB 121 KVIDTLTCGADLMGYIPLVGAAPLGGAAARLAHGVRLVLDGVNATGNLPGCSFIFLLA 180
 QY 181 LLSCLTIPASAVEVENVSGIVHVTNDCNSGIVYAAAVIMHTPCVPCVQEGNSRCWV 240
 DB 181 LLSCLTIPASAVEVENVSGIVHVTNDCNSGIVYAAAVIMHTPCVPCVQEGNSRCWV 240
 QY 241 ALTPTLAARNSSIPPTTIRRHVDLLVGAALCSAMVVDLCSVFLVQLTFSPRRYET 300
 DB 241 ALTPTLAARNSSIPPTTIRRHVDLLVGAALCSAMVVDLCSVFLVQLTFSPRRYET 300
 QY 301 VQDCNCSLYPGHVSCHRMAMDMNWSPTTALVVSQLLRIPOAVVDMVAGAHVGLAGLA 360
 DB 301 VQDCNCSLYPGHVSCHRMAMDMNWSPTTALVVSQLLRIPOAVVDMVAGAHVGLAGLA 360
 QY 361 YYSVMGNWAKVLIIVALLPAGVDGETHTTGRVAGHTTSGFTSIFSGASQKOLVNTNGSW 420
 DB 361 YYSVMGNWAKVLIIVALLPAGVDGETHTTGRVAGHTTSGFTSIFSGASQKOLVNTNGSW 420
 QY 421 HINRTALNCNDSLOTGFFAALFYAKFNSSGCPERMASCRPIDFAQGWGPIYTKPNSS 480
 DB 421 HINRTALNCNDSLOTGFFAALFYAKFNSSGCPERMASCRPIDFAQGWGPIYTKPNSS 480
 QY 481 DORPCYCHYAPRPGVVPASVOCGPVICTPSPVVGTTDSGVTTSWGENETDMLLN 540
 DB 481 DORPCYCHYAPRPGVVPASVOCGPVICTPSPVVGTTDSGVTTSWGENETDMLLN 540
 QY 541 NTRPQGNWFGCTWNNSGTGTCGGPPCNTGGVGNRTLICPTCFRKHPEATYTKCGSG 600
 DB 541 NTRPQGNWFGCTWNNSGTGTCGGPPCNTGGVGNRTLICPTCFRKHPEATYTKCGSG 600
 QY 601 PWLPRCLVDYPRKLWHYPCTLNFSIFKRVYVGVGEHRLNAACNWTGERCNLEDRS 660
 DB 601 PWLPRCLVDYPRKLWHYPCTLNFSIFKRVYVGVGEHRLNAACNWTGERCNLEDRS 660
 QY 661 ELSPLLSTTTEWQLPCAFPTLIPALSLGLIHLHONIVDVQVLYGVGSAPFVSFAIKWEYL 720
 DB 661 ELSPLLSTTTEWQLPCAFPTLIPALSLGLIHLHONIVDVQVLYGVGSAPFVSFAIKWEYL 720

QY 721 LFLLLADARVACACIWMMLLIAQEAALLENLVVINAASVAGAHGILSFLVFFCAAWYIKG 780
Db 721 LFLLLADARVACACIWMMLLIAQEAALLENLVVINAASVAGAHGILSFLVFFCAAWYIKG 780
QY 781 RLAPGAAYAFYGVWPLILLILLALPPRAYALDRMAASCGGAVLVGLVFTLSPYKYVFLT 840
Db 781 RLVPGAAYAFYGVWPLILLILLALPPRAYALDRMAASCGGAVLVGLVFTLSPYKYVFLT 840
QY 841 RLILWLOEFTTRAEAHMOMVWVPPNVGRGGRDAIILLTCAVHPELIPDIITKLLAILLGPAM 900
Db 841 RLILWLOEFTTRAEACIQLWVPPNVGRGGRDAIILLTCAVHSELIPDIITKILLAILLGPAM 900
QY 901 VZQAGLTRVPYFVRAOGLIPACMLVEKVAGHYVQVMFKLIGALTYVYVNHLPDRDWA 960
Db 901 VLQAGLTRVPYFVRAOGLIPACMLVEKVAGHYVQVMFKLIGALTYVYVNHLPDRDWA 960
QY 961 HAGLRDLAVAPVVPVFSANETKVIITWADTAACGGDIILGLPVSARRGKFIIFGPADSLEG 1020
Db 961 HSGLRDLAVAPVVPVFSANETKVIITWADTAACGGDIILGLPVSARRGKFIIFGPADSLEG 1020
QY 1021 QWRLLIAPITAYSQOTRGLVGCIIITSLGRDKNOVEGEVQVSTATOSFLATCINGVCWT 1080
Db 1021 QWRLLIAPITAYSQOTRGLVGCIIITSLGRDKNOVEGEVQVSTATOSFLATCINGVCWT 1080
QY 1081 VYHAGSKTLAGPKGFIINQMYTNVDQLVGMQAPGASALTPCTYGSBDLVIVTRHADVI 1140
Db 1081 VYHAGSKTLAGPKGFIINQMYTNVDQLVGMQAPGASALTPCTYGSBDLVIVTRHADVI 1140
QY 1141 PVRRGDSRGSLLSRPVSYLKSGGPIILCPSHGVVGFRAAVCTRGVAKAVDTIPVES 1200
Db 1141 PVRRGDSRGSLLSRPVSYLKSGGPIILCPSHGVVGFRAAVCTRGVAKAVDTIPVES 1200
QY 1201 METTMRSPVFTDNSTPPAVPQTFQVAHLHAPTGSCKSTKVPAAVAAQGVKVLVNLPSVAA 1260
Db 1201 METTMRSPVFTDNSTPPAVPQTFQVAHLHAPTGSCKSTKVPAAVAAQGVKVLVNLPSVAA 1260
QY 1261 TLGFGAYMSKAGHIDPNTIGVTRITIGSTYTYGKPLADGGCGGGAYDIIICDECHS 1320
Db 1261 TLGFGAYMSKAGHIDPNTIGVTRITIGSTYTYGKPLADGGCGGGAYDIIICDECHS 1320
QY 1321 TDSITLIGITVLDQAEATAGALVVLATATPGSVTVVPHNIEELGLSNBIEIPYKGA 1380
Db 1321 TDSITLIGITVLDQAEATAGALVVLATATPGSVTVVPHNIEELGLSNBIEIPYKGA 1380
QY 1381 PLEAIKGRHLIFCHSKKKDELAALKGLGINAVAYYRGLDVSVIPPIGVVWVATDAL 1440
Db 1381 PLEAIKGRHLIFCHSKKKDELAALKGLGINAVAYYRGLDVSVIPPIGVVWVATDAL 1440
QY 1441 MTGTFGDFSDVIDCNTCVTVQVDFSLDPTFTIETTVPODAVSRQGRGTGRGSGIYR 1500
Db 1441 MTGTFGDFSDVIDCNTCVTVQVDFSLDPTFTIETTVPODAVSRQGRGTGRGSGIYR 1500
QY 1501 FVTPGERESGMFDSVLCEDYDAGCAWYELTPEATSVELRAYLNTPGIPVQDHLIFWES 1560
Db 1501 FVTPGERESGMFDSVLCEDYDAGCAWYELTPEATSVELRAYLNTPGIPVQDHLIFWES 1560
QY 1561 VFTGLTHIDAHFLSOTKQAGDNFPLVAYOATVCAQAAPPSPDWOMKCLIRLXPTLHG 1620
Db 1561 VFTGLTHIDAHFLSOTKQAGDNFPLVAYOATVCAQAAPPSPDWOMKCLIRLXPTLHG 1620
QY 1621 PTPLLYRLGAVQNEVILTHPTIKYIMACMSADLEVVVSTWVWVGVLAALAAAYCLITGVS 1680
Db 1621 PTPLLYRLGAVQNEVILTHPTIKYIMACMSADLEVVVSTWVWVGVLAALAAAYCLITGVS 1680
QY 1681 VIVGRILLSGKPAVVPDREVLVQREDEBECASOLPYIEQCMQALAEQKQKALGLLQTAT 1740
Db 1681 VIVGRILLSGKPAVVPDREVLVQREDEBECASOLPYIEQCMQALAEQKQKALGLLQTAT 1740
QY 1741 KOAEAAAPVWESKWEALFTFWAKHWNPIISGIVLAGLSTLPGNPAIASIAFTASITSP 1800
Db 1741 KOAEAAAPVWESKWEALFTFWAKHWNPIISGIVLAGLSTLPGNPAIASIAFTASITSP 1800
QY 1801 LTTQNTLLFNILGGVAAQAAPPSSAASFVAGIAGAAVSGIGLKVLDVILAGYGAGVA 1860

Db 1801 LTTQNTLLFNILGGVAAQAAPPSSAASFVAGIAGAAVSGIGLKVLDVILAGYGAGVA 1860
QY 1861 GALVAFKVMSEVPSTEDLVNLLPAILSPGALVWVCAAILRRHVGPBGAVQWMMRLI 1920
Db 1861 GALVAFKVMSEVPSTEDLVNLLPAILSPGALVWVCAAILRRHVGPBGAVQWMMRLI 1920
QY 1921 APASGNHVSPTHYVPESDAARVTOILSSITITOLLKRLHQNEDCSTPCSGSWLRDV 1980
Db 1921 APASGNHVSPTHYVPESDAARVTOILSSITITOLLKRLHQNEDCSTPCSGSWLRDV 1980
QY 1981 MDWICTVLITDCKTLQSKLLPRLPGVPFLSCORGKYGWVRGDMQTTCPGCAQITAGHVK 2040
Db 1981 MDWICTVLITDCKTLQSKLLPRLPGVPFLSCORGKYGWVRGDMQTTCPGCAQITAGHVK 2040
QY 2041 NGSNRIVGPRTCSNTHGCTFPINAYTTGCTPSPAPNYSRALWRVAAEYEVTVRGDFH 2100
Db 2041 NGSNRIVGPRTCSNTHGCTFPINAYTTGCTPSPAPNYSRALWRVAAEYEVTVRGDFH 2100
QY 2101 YVTCGNTDNVPCQVPAPEFTEVDGVLHRYAPACKPLILREDDTVQVGLNQHLVGSOL 2160
Db 2101 YVTCGNTDNVPCQVPAPEFTEVDGVLHRYAPACKPLILREDDTVQVGLNQHLVGSOL 2160
QY 2161 PCEPDPVTVLTSLMTDPSHITATKARLARGSPSPSSASSASQLSAPSLSKATCTTHD 2220
Db 2161 PCEPDPVTVLTSLMTDPSHITATKARLARGSPSPSSASSASQLSAPSLSKATCTTHD 2220
QY 2221 SPDADLIEANLLWQEMGNITRVESENKVVILDSFEPLHABGDEREISVAEILKSRK 2280
Db 2221 SPDADLIEANLLWQEMGNITRVESENKVVILDSFEPLHABGDEREISVAEILKSRK 2280
QY 2281 PPSALPIWARCPYNNPPLILSKWDYVVPVWHGCELPPTKAPPIPPPRKKTIVLTESNV 2340
Db 2281 PPSALPIWARCPYNNPPLILSKWDYVVPVWHGCELPPTKAPPIPPPRKKTIVLTESNV 2340
QY 2341 SSALAEALATKTFGSSGSAVDSGTATLPLDASDDGKSDVESYSMMPLLEGPGDPL 2400
Db 2341 SSALAEALATKTFGSSGSAVDSGTATLPLDASDDGKSDVESYSMMPLLEGPGDPL 2400
QY 2401 SDGWSNVSBEASDDVCCMSYTWGTALITPCAAAEFSKLPINFLSNLRLHNNVYATT 2460
Db 2401 SDGWSNVSBEASDDVCCMSYTWGTALITPCAAAEFSKLPINFLSNLRLHNNVYATT 2460
QY 2461 SRSAGLRQKVTPLRQVLDHVDLKEKMAKASTVAKALLSVEEACKLTPPHSAKSF 2520
Db 2461 SRSAGLRQKVTPLRQVLDHVDLKEKMAKASTVAKALLSVEEACKLTPPHSAKSF 2520
QY 2521 GYGAKDVRLNLSRAVNHIRSVMEDLLEDTETPIDTTIMAKSEVFCVQPKGRKPRLIV 2580
Db 2521 GYGAKDVRLNLSRAVNHIRSVMEDLLEDTETPIDTTIMAKSEVFCVQPKGRKPRLIV 2580
QY 2581 FPDILGVRYCEKMAIYDVVSTLPQAVMGSSYGFQVSPQRVEFLVNTWKSCKCPMFSYDT 2640
Db 2581 FPDILGVRYCEKMAIYDVVSTLPQAVMGSSYGFQVSPQRVEFLVNTWKSCKCPMFSYDT 2640
QY 2641 RCFDSTVTESDIRVEBSIYOCCLDAPARQAIJSLTERLXIGGPLETMSKQNCYRRCRA 2700
Db 2641 RCFDSTVTESDIRVEBSIYOCCLDAPARQAIJSLTERLXIGGPLETMSKQNCYRRCRA 2700
QY 2701 SGVLTTCGNTLTCYLKATAACRAAKLQDCTMLVNGDDLVIICESAGTQDAALRAFTE 2760
Db 2701 SGVLTTCGNTLTCYLKATAACRAAKLQDCTMLVNGDDLVIICESAGTQDAALRAFTE 2760
QY 2761 ANTRYSAPPDPPQPEYDLELITSCSNVSVADHAGKRVYVYLTROPTTFLAANAWEAR 2820
Db 2761 ANTRYSAPPDPPQPEYDLELITSCSNVSVADHAGKRVYVYLTROPTTFLAANAWEAR 2820
QY 2821 HTPINSMLGNIIMYAPTLWARMILMTHFTSIIIAQOELEKALCOIYAGCISTEPDLPO 2880
Db 2821 HTPINSMLGNIIMYAPTLWARMILMTHFTSIIIAQOELEKALCOIYAGCISTEPDLPO 2880
QY 2881 IIBRLHGLSAFTLHYSVPGINEVASCRLKGLVPPPLTWHRHARSVRAKLLSOGGRAATC 2940

```

Db      2881  IIQRLHGLSFLSHSYSPGEINRVASCLRLKLVPEPLVRHRARSVRKLLSQCGRAATC 2940
Qy      2841  GRYLFNNAVRKTLKLTPIPAASQDLSCWFFVAGYSGGDIYHLSGRARWPLCLILLVS 3000
Db      2841  GRYLFNNAVRKTLKLTPIPAASQDLSCWFFVAGYSGGDIYHLSGRARWPLCLILLVS 3000
Qy      3001  GVGIYLLFNR 3010
Db      3001  GVGIYLLFNR 3010

RESULT 6
GNWVC3
Genome polyprotein - hepatitis C virus (strain HCV-1)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HCV-1) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C;Accession: A39166; PQ0403; PQ0404
R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi
Proc. Natl. Acad. Sci. U.S.A. 89, 2451-2455, 1991
A;Title: Genetic organization and diversity of the hepatitis C virus.
A;Reference number: A39166; MUID:91172826; PMID:1848704
A;Accession: A39166
A;Molecule type: mRNA
A;Residues: 1-3011 <CHO>
A;Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to a
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0403
A;Molecule type: Genomic RNA
A;Residues: 1577-1633 <CHA>
A;Cross-references: DDBJ:D10128
A;Experimental source: Isolates E-b16
A;Accession: PQ0404
A;Molecule type: Genomic RNA
A;Residues: 1577-1633 <CH2>
A;Experimental source: Isolates E-b17
A;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <WEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus predicted <NS3>
F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22
Query Match 88.2%; Score 14119.5; DB 1; Length 3011;
Best Local Similarity 85.2%; Pred. No. 0;
Matches 2564; Conservative 226; Mismatches 220; Indels 1; Gaps 1;

Qy      1  MSTNPKPQRTKNTNRPDQVFPGGQIVGGYLLPRRGPRLGVRAATRKASERSQPRG 60
Db      1  MSTNPKPQKNTNRPDQVFPGGQIVGGYLLPRRGPRLGVRAATRKASERSQPRG 60
Qy      61  RROIPKARPEGRAPQPCWPELYNGLWAGWLLSRGRSPSWGPTDPRRSENIG 120
Db      61  RROIPKARPEGRAPQPCWPELYNGLWAGWLLSRGRSPSWGPTDPRRSENIG 120
Qy      121  KVITDLTCGFADLMGYIPLVGAFLGGARALAHGVVRVLEQVNVATGNLPGCSFSLPLA 180
Db      121  KVITDLTCGFADLMGYIPLVGAFLGGARALAHGVVRVLEQVNVATGNLPGCSFSLPLA 180

```

```

Qy      181  LLSCLTIPASAYEVNRVSGIYHVTNDCSSSIYVEAADVIMHTPGCVPCVQEGNSRCWV 240
Db      181  LLSCLTIPASAYQVRNSTGLYHVTNDCPNSSIIYVEAADAILHTPGCVPCVQEGNASRCW 240
Qy      241  ALPTPTLAARVASVPTTIRRHVDLLVGTAAFCAMVVDLCGSIPLVSQLTFTSPRHET 300
Db      241  AMPTVATRDGKLPAQLRRHIDLLVGSATLCSALYVDLCGSAFLVGQLTFTSPRHWT 300
Qy      301  VQDCNSIYPCHVSGHRMAWDMAMNSPTTALVVSQLLRIPQAVVDVAGAHGVLAGLA 360
Db      301  TQGCNSIYPCHITGHRMAWDMAMNSPTTALVMAQLLRIPQAILDIAEGAENVLAGIA 360
Qy      361  YSMVGNWAKVLVALLFAGVDGTHTCRVAGHTTSGFTSLSPSSASQKIQLVNTNGSW 420
Db      361  YFSVGNWAKVLVLLFAGVDATHTVGTGSAGHTVSGFVSLAPGAKQXVQINTNGSW 420
Qy      421  HINRTALNCDSLQGTFFAALFYAHKFNSSGCPERMASCRPIDWFAQGGWPIYTKPNSS 480
Db      421  HLNSTALNCNLSLNTGWLGLFYVHKFNSSGCPERLASCEPLTDFDQGWGPISYANGSGP 480
Qy      481  DQRPYCHWYAPRCGVVPASQVCGPVYCTPSPVYVGTTRDSRGVPTYSNGENETDMLN 540
Db      481  DQRPYCHWYPPKPCGIVPAKSVCGPVYCTPSPVYVGTTRDSRGVPTYSNGENDTDFVLN 540
Qy      541  NTRPPQGNWFGCTWMSNSTFTKTCGGPPCNIGVGNRTLICPTDCPKHPEATYTKCGSG 600
Db      541  NTRPPQGNWFGCTWMSNSTFTKTCGGPPCNIGVGNRTLICPTDCPKHPEATYTKCGSG 600
Qy      601  PWLTPCLVDYPYRLWHEYFCTLNFSIFKVRMYVGGVEHRINAACNWTGRCCNLEDRDS 660
Db      601  PWLTPCLVDYPYRLWHEYFCTLNFSIFKVRMYVGGVEHRLEAECNWTGRCCNLEDRDS 660
Qy      661  ELSPLLLSTETEWQILPCAFITLTPALSTGLIHLHONIVDQVLYGVGSFAFVSPAIKWEYL 720
Db      661  ELSPLLLTTTQWQVLPCEFTLTPALSTGLIHLHONIVDQVLYGVGSFIASWAIKWEYV 720
Qy      721  LLEFLLADARVCACLWMLLIQAABALENLVNLNAAVAGAHGILSLVFFCAAAWYIKG 780
Db      721  LLEFLLADARVCSCLWMLLIQAABALENLVNLNAAVAGAHGILSLVFFCAAWYIKG 780
Qy      781  RLAPGAAYAFYGVWPLLLLLALLPPRAYALDREMAASCAGVAVGLVFTLSPYKVFELT 840
Db      781  KWVPGAAYTFYGVWPLLLLLALLPPRAYALDREMAASCAGVAVGLVGLMAULTLSPYKRYIS 840
Qy      841  RLHWLQYFRTAREAWQVWVPLNVRGGRDAIILLTCAVHPHLLIEDIKLLAILGLPLM 900
Db      841  WCLWMLQYFRTAREAWQVWVPLNVRGGRDAIILLTCAVHPHLLIEDIKLLAILGLPLW 900
Qy      901  VLOAGITRVPYFVYVQAQGLIRACMLVRKAVGSHVQVWFMKLGALTCTYVYNHLLTPRDWA 960
Db      901  ILQASLLKVPYFVYVQAQGLIRACMLVRKAVGSHVQVWFMKLGALTCTYVYNHLLTPRDWA 960
Qy      961  HAGRLAVABEVPVPSAMETKVTWGADTAACGDIILGLPVSAARGKEIFIGPADSLRG 1020
Db      961  HNGRLAVABEVPVPSAMETKVTWGADTAACGDIILGLPVSAARGKEIFIGPADGWS 1020
Qy      1021  QGRELAPATAYSOQTRGVGLCIIITSLTCRDNKNQVEGYQVYSTATQSLFATCINGCVHT 1080
Db      1021  KGRELAPATAYSOQTRGVGLCIIITSLTCRDNKNQVEGYQVYSTATQSLFATCINGCVHT 1080
Qy      1081  VYHAGSKTLAPKGPITQWMTNVDLDLVGWAOPCGARSMTPCSCGSSDLVLTWHADVI 1140
Db      1081  VYHAGTRIASPKGPVIOWMTNVDLDLVGWAOPCGARSMTPCSCGSSDLVLTWHADVI 1140
Qy      1141  PYRRGDSRGSLLSPRPISYLXGSGGPIICPSGHVVGVRFAAVCTRGVAKAVDFIPBS 1200
Db      1141  PYRRGDSRGSLLSPRPISYLXGSGGPIICPSGHVVGVRFAAVCTRGVAKAVDFIPVEN 1200
Qy      1201  METTMSPEVTNSTPPAVPQTFVAHLHAPTGSGKSTKVPAAQAQGYKVLVNPVAA 1260
Db      1201  LETTMSPEVTNSTPPAVPQTFVAHLHAPTGSGKSTKVPAAQAQGYKVLVNPVAA 1260
Qy      1261  TLGFGAYMSKAGIDPNIRTGVRTITTTGGSITVSTYCKFLADGGCGGGAYDIIICDECHS 1320

```


F;2-115/Product: capsid protein C #status predicted <CPC>
 F;116-191/Product: envelope protein M #status predicted <EPW>
 F;192-389/Product: major envelope protein E #status predicted <VE>
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F;1007-1615/Product: hepatitis virus #status predicted <NS3>
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)
 F;1312-1317/Region: nucleotide-binding motif B
 F;1316-1319/Region: DEXH motif
 F;1616-1616/Product: nonstructural protein NS4a #status predicted <NS4a>
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 87.5%; Score 14008.5; DB 1; Length 3011;
 Best Local Similarity 84.9%; Pred. No. 0;
 Matches 2555; Conservative 219; Mismatches 236; Indels 1; Gaps 1;

QY	1	MSNTPQKTKNTNRPODVPFPGGGQIVGVYLLPRGPRLGVRATKASERSQPRG	60
DB	1	MSTIPKQKTKNTNRPODVPFPGGGQIVGVYLLPRGPRLGVRATKASERSQPRG	60
QY	61	RRQPIKARPEGRWAQPGYFWPLYGNEGLWAGWLLSPRGSPSGMGPTDPRRSRLG	120
DB	61	RRQPIKARPEGRWAQPGYFWPLYGNEGLWAGWLLSPRGSPSGMGPTDPRRSRLG	120
QY	121	KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNATGNLPGCSFSIFLLA	180
DB	121	KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNATGNLPGCSFSIFLLA	180
QY	181	LLSCLIPASAYVRNVSGIHYHNTDCNSISYVEAADVMHTPGCVPCVQEGNSRCWV	240
DB	181	LLSCLVPASAYVRNVSGIHYHNTDCNSISYVEAADVMHTPGCVPCVQEGNSRCWV	240
QY	241	ALPTETLAARNASVPTTIRHVDLLVGTAAFCAMVVGDLGSGSIFLVSQLFTSPSPRHET	300
DB	241	AMPTVATRDGKLPATQRRHIDLLVGSATLCSALYVGLDGGSVFLGQFTFSPRRHT	300
QY	301	VQDCNSIYPGHVSGHRMAMDMNNSPTTALVVSQLLRIPOAVDMVAGAHGVLGAGLA	360
DB	301	TQGCNSIYPGHTTGRMAMDMNNSPTTAAALVMAQLLRIPQAILDMIAHAGWVLGAGLA	360
QY	361	YISMVGNWAKVLVALLPAGVDGETHTGRVAGHTVSGFTSLFSSGASQKIQLVNTNGSW	420
DB	361	YISMVGNWAKVLVLLFAGVDDETIVSGQAAAMSGVLSLFTPGAKQNIQLINTNGSW	420
QY	421	HINRTALNCDSLOTGFPAALFAHAFNSGSCPERMASCRPIDWFAQGWGPIYTRPNS	480
DB	421	HINSTALNCNESLNTGLAGLIYQHKFNSSGCPERLASCRJLTDFOGWGPISHANGSGP	480
QY	481	DQRPYCNHVAAPRCGVVPASQCVGVYCFPTSPVVGTTDRSGVPTYSWGENETDYMILN	540
DB	481	DQRPYCNHVPKPGIIPAKSVGCGPVYCFPTSPVVGTTDRSGAPVYVNWGANDTDVFLN	540
QY	541	NTRPPQGNWFCGWMNSTGFTKCGGPPCNIGVGNRTLCPTDCPRKIPKPAATYKCGSG	600
DB	541	NTRPPQGNWFCGWMNSTGFTKCGGPPCNIGVGNRTLCPTDCPRKIPKPAATYKCGSG	600
QY	601	PWLTPLCLVDYPRYLWHYPCITLNFSPKVMYVGGVEHRLNAACNWRGRCMLNDRDS	660
DB	601	PWLTPLCLVDYPRYLWHYPCITNITFKVMYVGGVEHRLDAACNWRGRCMLNDRDS	660
QY	661	ELSPILLSSTTEWQILPCAFPTTLPALSTGLIHLHQNIVDVQYLVGVSAFVSFAIKWEYIL	720
DB	661	ELSPILLSSTTQVQLPFCSTTLPALSTGLIHLHQNIVDVQYLVGVSSIASMAIKWEYV	720
QY	721	LLPFLILLADARVCAACLMWLLIAQAEALENLVINAASVAGAHGILSLVFFPCAANYIKG	780
DB	721	LLPFLILLADARVCSCLMWLLISQAEALENLVINAASAGTGLVSLVFFCFAWYILKG	780
QY	781	RLAFGAAYAFYGVWPLILLILLALLPPRAYALDRWAASCGGAVLVGLVFTLSFYKVFELT	840
DB	781	RWVPGAAYALYGNWPLILLILLALLPQRAYALDTEVAASCGGVVLGLMALTSLSYFRCIS	840

QY	841	RLIMWLOXFITRAEAHMOVVPPPLNVGRGRDAIILLTCAVHPFLIFDITKLLAILGFLM	900
DB	841	WCLMWLOFTRAEQAHLVWVPPPLNVGRGRDAVILLMCMVHTFLVFDITKLLAVLGLM	900
QY	901	VLQAGITRVPVPRVQAQGLIRACNLVRKAGHYQVMFKGALITGTIVYVNHLLTPIROWA	960
DB	901	ILQASLLKVPYFVRVQGLIRICALARKMVGHYQVMAILKLGALTGTIVYVNHLLTPIROWA	960
QY	961	HAGLRDLAVAVEPVVFSAMETKVIITWAGDTAACGDIILGLPVARSRRGEIIFLGSPADSLEG	1020
DB	961	HNLGLDLAVAVEPVVFSQWETKLIITWAGDTAACGDIINGLPVARSRRGEIILGSPADGWS	1020
QY	1021	QGWRLLPATITAYSQQTRGVLCIITSLTRDKNQVEGEVQVVSSTATQFLATCINGVCWT	1080
DB	1021	KWRLLPATITAYAQQTRGLLGCIITSLTRDKNQVEGEVQIVSTAAQFLATCINGVCWT	1080
QY	1081	VYHAGSKTLAPGKPIOMYTNVDLDIVGWQAPPGARSMTPCSCGSDLYLVHRAHVI	1140
DB	1081	VYHAGTTLTASPKGPVIQMYTNVDQDIVGWPAQPGARSMTCTCGSSDLYLVHRAHVI	1140
QY	1141	PVRRGDSRGLSLSPRPVSYLKGSGGGLLCPSHVGVGFRAAVCTRGVAKAVDFIPVES	1200
DB	1141	PVRRGDSRGLSLSPRPISYLYLKGSGGGLLCPAGHVGLFRAAVCTRGVAKAVDFIPVES	1200
QY	1201	METMRSVFTDNSTPPAVPQTFQVAHLHAPTSGCKSTKVPAAVAAQGYKVLVNLPSVAA	1260
DB	1201	LETTMRSVFTDNSTPPAVPQSFQVAHLHAPTSGCKSTKVPAAVAAQGYKVLVNLPSVAA	1260
QY	1261	TLGFGAYMSKAHGDIPNIRTGRTITTCGSIITYSYGKFLADGGCGGAYDIIICDECHS	1320
DB	1261	TLGFGAYMSKAHGDIPNIRTGRTITTCGSIITYSYGKFLADGGCGGAYDIIICDECHS	1320
QY	1321	TDSTTILGIGTVLQDAETAGARLVLAATATPPGVTVEPHNIEHIGLNNGBIPYGRKAI	1380
DB	1321	TDATSVLIGTVLQDAETAGARLVLAATATPPGVTVEPHNIEHIGLNNGBIPYGRKAI	1380
QY	1381	PIEAIKGRHILFCHSKKCDLAAKLTGLGLNNAVAYYRGDLVSVIPIPGVWVATDAL	1440
DB	1381	PLEAIKGRHILFCHSKKCDLAAKLVAGLVNAVAYYRGDLVSVIPIPGVWVATDAL	1440
QY	1441	MTGFTGDDSDVLDNCTVQTVDRSLDPTFIETTTVQDASVRSORGRTRGSGSYR	1500
DB	1441	MTGYTGDDSDVLDNCTVQTVDRSLDPTFIETTTLPQDASVRSORGRTRGSGSYR	1500
QY	1501	FVTPGERSGMDSSVLCEVDAGCAWYSLTPARTSVRLRAYLNTPLGVPVCHDLFEWES	1560
DB	1501	FVAPGERSGMDSSVLCEYDTGCANVELTPEATTVALRAYVNTPLGVPVCHDLFEWES	1560
QY	1561	VFTGLTHIDAHFLSQTQKAGDNFPYLVAQATVCARAQAPPPSWDMKCLIRLKPTLHG	1620
DB	1561	VFTGLTHIDAHFLSQTQKAGDNFPYLVAQATVCARAQAPPPSWDMKCLIRLKPTLHG	1620
QY	1621	PTPLAYRLGAVQNEVLTHTPIKYIACMSADLEVVTSTWLVGVLAALAAAYCLTTSV	1680
DB	1621	PTPLAYRLGAVQNEVLTHTPIKYIMTMSADLEVVTSTWLVGVLAALAAAYCLTTSV	1680
QY	1681	VIVGRILSLSGKPAVVDREVLVYQEEDEMEECASQLEPIEQGMQLAEQFKQKALGLQAT	1740
DB	1681	VIVGRILSLSGFAIIPDREVLVYQEEDEMEECASQLEPIEQGMQLAEQFKQKALGLQAT	1740
QY	1741	KQABAAAAPVWESKWALETFWAKHKNWFIISGLQYLAGLSTLPGNPAIASLMAFTASISP	1800
DB	1741	ROAEVIAPTVNTWOKLEAFWAKHKNWFIISGLQYLAGLSTLPGNPAIASLMAFTASISP	1800
QY	1801	LTTQNTLPLNLLGVAQALAPPSAASAFVAGIAGAAVGSIGLGLVLDIILAGVAGVA	1860
DB	1801	LTTQNTLPLNLLGVAQALAPPSAASAFVAGIAGAAVGSIGLGLVLDIILAGVAGVA	1860
QY	1861	GALVAFKMSGVPSTEDLVNLLPAILSPGALVWGVCAAILRRHVGPGEQAVQWNNELI	1920
DB	1861	GALVAFKMSGVPSTEDLVNLLPAILSPGALVWGVCAAILRRHVGPGEQAVQWNNELI	1920
QY	1921	AFASRGNHVSPTHYVPSDAAARVQIILSSLTITQLKKLHQWINEDCSTPCSGSNLRDV	1980

Db 1921 AFASRGNHVSPTFFVPPSDAAARVTAISSLTIVTQLLRLHQMLSSBSTTTCGSGWLRDI 1980
Qy 1981 WDMICTVLTDPKTKWLOSKLPLRPGVFLSCORGKGVKWRGDTMQTCTCGQAIAGHYK 2040
Db 1981 WDMICEVLSDFKTKWTKLMPHDFGIFBVSCHQYKGVKWRGDTMQTCTCGQAIAGHYK 2040
Qy 2041 NGSMRIVGPRTCSNTHGTTPPINAYTGTCTPPSPAPNYSRALMEVAEEYVVRVGDHF 2100
Db 2041 NGTMRIVGPRTCSNTHGTTPPINAYTGTCTPLPAPNYTALMEVSAEYVVRVGDHF 2100
Qy 2101 YVGTMTDNKCPQVPAPFFTEVDGVRHLRHVAPACKPILLRDTVFQVGLNQVLSQSL 2160
Db 2101 YVGTMTDNKCPQVPAPFFTEVDGVRHLRHVAPACKPILLRDTVFQVGLNQVLSQSL 2160
Qy 2161 PCPEPDPVTLTSMLTDPSTHTATKRRRLARGSPPSLASSASOLGAPSLKATCTTHD 2220
Db 2161 PCPEPDPVTLTSMLTDPSTHTATAAAGRLARGSPPSLASSASOLGAPSLKATCTTHD 2220
Qy 2221 SPDAELIENLLWRQEMGNITRVESNKKVILDSFFPLHAEGDEREISVAAELIRKSRK 2280
Db 2221 SPDAELIENLLWRQEMGNITRVESNKKVILDSFDPPLVAEEDEREISVPAELIRKSR 2280
Qy 2281 FPSALPIWARPDPNPLLESKDPDYPPVHGCPLPPTKAPPIPPRRKRTVVLTESNV 2340
Db 2281 FTQALPIWARPDPNPLLETWKENYBPPVHGCPLPPTKAPPIPPRRKRTVVLTESNV 2340
Qy 2341 SSALARIATKTFGSGGSANDGTATLPOLASDDGKGSVSEYSPMPLEGGPDPDL 2400
Db 2341 STALAEIENLLWRQEMGNITRVESNKKVILDSFDPPLVAEEDEREISVPAELIRKSR 2400
Qy 2401 SDGWSVTSSEA-SEDYVCCSMSTWTGALITPCAAEESKIPINPLNSLIRHNNMYAT 2459
Db 2401 SDGWSVTSSEA-SEDYVCCSMSTWTGALITPCAAEESKIPINPLNSLIRHNNMYAT 2459
Qy 2460 TSRSASIRQKVTEDRLQVLDHVDVLEKMAKASTVKAALLSIEACKLTPPHSKSK 2519
Db 2461 TSRSASIRQKVTEDRLQVLDHVDVLEKMAKASTVKAALLSIEACKLTPPHSKSK 2519
Qy 2520 FGVGAKDVRNLSRAVNIHRSWEDLLEDTPTDITIMAKSEYVQVPEKGGKRPALRI 2579
Db 2521 FGVGAKDVRNLSRAVNIHRSWEDLLEDTPTDITIMAKSEYVQVPEKGGKRPALRI 2579
Qy 2580 VFPDLGVRCERKALYDVWSTLPOAVNGSSYGFQYSPQRVFEVLNWKSKKCPMGFSYD 2639
Db 2581 VFPDLGVRCERKALYDVWSTLPOAVNGSSYGFQYSPQRVFEVLNWKSKKCPMGFSYD 2639
Qy 2640 TRCPDSTVTSDDIRVESIYOCCLDAPARQAIISLTERLYIGGPIIENSKQNGCYRRC 2699
Db 2641 TRCPDSTVTSDDIRVESIYOCCLDAPARQAIISLTERLYIGGPIIENSKQNGCYRRC 2699
Qy 2700 ASGVLTTCGNTLCYKATACAAKALQDCTMLVNGDDLVIICESAGTOEDAAALRAFT 2759
Db 2701 ASGVLTTCGNTLCYKATACAAKALQDCTMLVNGDDLVIICESAGTOEDAAALRAFT 2759
Qy 2760 EAMTRYGAPDPPQPDYDELITSCSNVSVVAHDASGRVYLYTRDPTTPLAANAETA 2819
Db 2761 EAMTRYGAPDPPQPDYDELITSCSNVSVVAHDGTCRVYLYTRDPTTPLAANAETA 2819
Qy 2820 RETPINSWLNMIINWAPTLNARMILMTHFFSILLAQBOLEKALDQIYACYSIEPDLIP 2879
Db 2821 RHTPVNSWLNMIINWAPTLNARMILMTHFFSILLAQBOLEKALDQIYACYSIEPDLIP 2879
Qy 2880 QIIERLHGLSAFTLHSPGSEINRVASCLRLKGLGVPPLKRWHRARSVRKLLSOGGRAAT 2939
Db 2881 PIIQELHGLSAFTLHSPGSEINRVAACLRKLGVPPLKRWHRARSVRKLLSOGGRAAT 2939
Qy 2940 CGRYLFENWAVEKTLKLPFAASQDLISGWTFVAGYSGGDIYHLSRARPRFELCLLIS 2999
Db 2941 CGRYLFENWAVEKTLKLPFAAAGRLISGWTFVAGYSGGDIYHLSRARPRFELCLLIS 2999
Qy 3000 VGVGCIYLLPNR 3010
Db 3001 AGVGIYLLPNR 3011

RESULT 6

genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A36814; A1546
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>
A;Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: Ap; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <BPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;156,209,234,305,335,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match: 97.0%; Score 13935.5; DB 1; Length 3011;
Best Local Similarity 84.5%; Pred. No. 0;
Matches 2545; Conservative 226; Mismatches 239; Indels 1; Gaps 1;

Qy 1 MSTNPKPQRTKNTNRPODVKPPGGQIVGGVYLLPRGPRLVGVRATKASRSQPRG 60
Db 1 MSTNPKPQRTKNTNRPODVKPPGGQIVGGVYLLPRGPRLVGVRATKASRSQPRG 60
Qy 61 RRQPIKARPEGRWAQPGYPMPLYNCGELGWAGWLLSPRGRPSWGPTDPRRRNLG 120
Db 61 RRQPIKARPEGRWAQPGYPMPLYNCGELGWAGWLLSPRGRPSWGPTDPRRRNLG 120
Qy 121 KVITDTTCGPADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNIPCCGFSIFLLA 180
Db 121 KVITDTTCGPADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNIPCCGFSIFLLA 180
Qy 181 LLSCLATPASAYEVNVSIGYHYVNTDCNSNIIYVAAADVTMTHTPGCVPCVQEGNSRRCW 240
Db 181 LLSCLATPASAYEVNVSIGYHYVNTDCNSNIIYVAAADVTMTHTPGCVPCVQEGNSRRCW 240
Qy 241 ALTPTLAARNASVPTTIRHVDLLVGTAAFCAMVYVGLCGSIFLVSQLFTFSPRRHET 300
Db 241 AVPTVATRGKLPFTQLRHHILLVGSATLCSALYVGLDLCGSLVGLQFTFSPRRHET 300
Qy 301 VQDCNSIYPGHVSCHVRAWMNNSPTTALVVSQLLRIPQAVVDMVAHGVLAGLA 360
Db 301 TQDCNSIYPGHVSCHVRAWMNNSPTTALVVSQLLRIPQAVVDMVAHGVLAGLA 360
Qy 361 YYSVGNWAKVLLVAILFAGVQGETHTTGRVAGHTTSGFTSLFSSGASQIKLVNTNGSW 420
Db 361 YYSVGNWAKVLLVAILFAGVQGETHTTGRVAGHTTSGFTSLFSSGASQIKLVNTNGSW 420

421 HINRTALNCDSJQTCFFAALFYAHKFNSSGCCPERMASORPIDWFAQGCPITYTKPNSS 480
421 HINRTALNCDSJQTCFFAALFYAHKFNSSGCCPERMASORPIDWFAQGCPITYTKPNSS 480
481 DORPYCWHYAPRCGVVPASQVGPVYCTPTSPVVGTTDRSGVPTYSGENETDWMLLN 540
481 DESPYCWHYAPRCGVVPASQVGPVYCTPTSPVVGTTDRSGVPTYSGENETDWMLLN 540
541 NTRPPGNNFGCTWMSNTOFTKTCGGPPCNIGVGNRTILCPTDCRKHPEATYTKCGSG 600
541 NTRPPGNNFGCTWMSNTOFTKTCGGPPCNIGVGNRTILCPTDCRKHPEATYTKCGSG 600
601 PMTPTCLVDYPRYLWHPCTNLNFSFKVRMYGVGVEHRLNAACNTRGERCNLEDRDRS 660
601 PRITPRCWDYPRYLWHPCTNLNFSFKVRMYGVGVEHRLNAACNTRGERCNLEDRDRS 660
661 EUSPLLLSTTEWOLPCAFMTLLPALSTGLHLHONIVDOYLTVGVGSAPVSPFAIKWEYL 720
661 EUSPLLLSTTEWOLPCAFMTLLPALSTGLHLHONIVDOYLTVGVGSAPVSPFAIKWEYL 720
721 LFLLLADARVACIWMMLLIAOAEALENLVNDAASVAGAHGILSFVFFCAANYIKG 780
721 LFLLLADARVACIWMMLLIAOAEALENLVNDAASVAGAHGILSFVFFCAANYIKG 780
781 RLAPGAAYAFYGVWPLLLILAILPPRAYALDRMAASCGGAVLVGLVFLTSLSPYKVFELT 840
781 RWVPGAVVLYGWPLLLILAILPPRAYALDRMAASCGGAVLVGLVFLTSLSPYKVFELT 840
841 RLIIWNYQFIFRABABQWVPPINVRGGRDAIILTCVHVPBELIDITKLLILAIJPLM 900
841 WCMWWLYQFLTRVEAOLHVVWPPINVRGGRDAIILTCVHVPBELIDITKLLILAIJPLM 900
901 VLOAGITRVPYFVRAOGLIRACMLVRKACGHVYVQVFMKLGALTGTVYVNHLPTRDWA 960
901 ILQASLKUPYFVRVQGLLICALARKIAGHVVQMAIKIGALTGTCTVNHLPTRDWA 960
961 HAGRLDAVAEPVFSAMETKVITWGADTAACGDIILGLPVSARRGKBIFFLPADSLG 1020
961 HNGRLDAVAEPVFSAMETKVITWGADTAACGDIILGLPVSARRGKBIFFLPADSLG 1020
1021 QGWELLAPITAYSOQTRGVJGCIITSLTGDRKNQVEGEVQVWSTATQSFATCINGVCWT 1080
1021 KGWELLAPITAYSOQTRGVJGCIITSLTGDRKNQVEGEVQVWSTATQSFATCINGVCWT 1080
1081 VYHAGSKTLGAPKPTQMYNVNDLQVQWAPPQARSMTPCSGSSDLYLVTRADVI 1140
1081 VYHAGSKTLGAPKPTQMYNVNDLQVQWAPPQARSMTPCSGSSDLYLVTRADVI 1140
1141 FVRRGDSRGLLSRPRVSYLKGSSGGPLLCPSGHVGVVGFRAAVCTRGVAKAVDFIPVES 1200
1141 FVRRGDSRGLLSRPRVSYLKGSSGGPLLCPSGHVGVVGFRAAVCTRGVAKAVDFIPVES 1200
1201 METTMSRPTDNTSTPAVQTFQVAHLHAPTSGSKTKVPAAYAGYKVLYNPBSVAA 1260
1201 LETTMSRPTDNTSTPAVQTFQVAHLHAPTSGSKTKVPAAYAGYKVLYNPBSVAA 1260
1261 TLGFGAYMSKAHGDIPNIRGVRTITITGGGITYSTYKFLADGCGSGAYDIIICDECHS 1320
1261 TLGFGAYMSKAHGDIPNIRGVRTITITGGGITYSTYKFLADGCGSGAYDIIICDECHS 1320
1321 TDSTTILGIVTLDOAFTAGARIVLATAPTPGSVTVPHNIBEIGLSNKGRIPTFYGKAI 1380
1321 TDATSIIGITVLDQAFTAGARIVLATAPTPGSVTVPHNIBEIGLSNKGRIPTFYGKAI 1380
1381 PIRAIKGRHLIPCHSKKKCDELAALKITGLNNAVAYRGVDYSVTPPIGVVWVATDAL 1440
1381 PLEVIKGRHLIPCHSKKKCDELAALKITGLNNAVAYRGVDYSVTPPIGVVWVATDAL 1440
1441 MTGFTGDFSDVDCNCTCVTQVDFSLDPTFTTETTTVPQAVRSORRGRTGRGSGIYR 1500
1441 MTGFTGDFSDVDCNCTCVTQVDFSLDPTFTTETTTVPQAVRSORRGRTGRGSGIYR 1500

1501 FVTPGERSPGMPDSSVLCEDAGCAWYELTPAETSVELRAYLNTPLGPVCDHLEFWES 1560
1501 FVTPGERSPGMPDSSVLCEDAGCAWYELTPAETSVELRAYLNTPLGPVCDHLEFWES 1560
1561 VFTGLTHIDAHPLSOTKQAGNFPYLVAIQATVCARAOPPPSWDMKMLRLRLKPTLHG 1620
1561 VFTGLTHIDAHPLSOTKQAGNFPYLVAIQATVCARAOPPPSWDMKMLRLRLKPTLHG 1620
1621 PPTILAYRGAVONEVILTHITKYIMACWSADLEVVTTSTWLVGGVLAALAAAYCIUTGSV 1680
1621 PPTILAYRGAVONEVILTHITKYIMACWSADLEVVTTSTWLVGGVLAALAAAYCIUTGSV 1680
1681 VIVGRILISGKPAVVPDREVLYQEFDEMEBCASQLPYIEQGNQLABQFKQKALGLQAT 1740
1681 VIVGRILISGKPAVVPDREVLYQEFDEMEBCASQLPYIEQGNQLABQFKQKALGLQAT 1740
1741 KOAEAAAAPVBSKWRALSTFWAKHWNFIISGTOYLAGLSTPLGNPAIASLMAFTASITSP 1800
1741 RHAEVITPAVQTNWQKLEVFNAKEWNFISGTOYLAGLSTPLGNPAIASLMAFTASITSP 1800
1801 LTTQNTLLFNILSGWVAQAAPPASAASAFVAGIAGAAGVSGIGLKVLDIILAGYGAVA 1860
1801 LTTQNTLLFNILSGWVAQAAPPASAASAFVAGIAGAAGVSGIGLKVLDIILAGYGAVA 1860
1861 GALVAFKMSGEVPTSTEDLVNLLPALISPGALVGVWCAAILRRRHVGPCEGAVOMNRLI 1920
1861 GALVAFKMSGEVPTSTEDLVNLLPALISPGALVGVWCAAILRRRHVGPCEGAVOMNRLI 1920
1921 AFASGNHVSPTHYVPESDAAARVTOILSSLTITOLLKELHOMINEDCSTPCSGSMLRDV 1980
1921 AFASGNHVSPTHYVPESDAAARVTOILSSLTITOLLKELHOMINEDCSTPCSGSMLRDV 1980
1981 WDWICTVITDFTKTLQSKLLPRLPGVPFLSCORGKGVWGRGDGIMQITCPCGAQIAGHYK 2040
1981 WDWICEVLSDFKTLQSKLLPRLPGVPFLSCORGKGVWGRGDGIMQITCPCGAQIAGHYK 2040
2041 NSGMRIVGPRTCNTHWGTTPNAYTTGCTPSPAPNTSRALWRVAABEYVEVTRVGDFH 2100
2041 NGTMRIVGPRTCNTHWGTTPNAYTTGCTPSPAPNTSRALWRVAABEYVEVTRVGDFH 2100
2101 YVTGMTDNNKPCQVAPAPETPEVDGVLRLHYAPACKPLLRDVTFOVLNOYLAVGSL 2160
2101 YVTGMTDNNKPCQVAPAPETPEVDGVLRLHYAPACKPLLRDVTFOVLNOYLAVGSL 2160
2161 POEPPDVTVLTSMLTDPSSHITAEAKRLARGSPPSLASSASQLSAPSKATCTTHD 2220
2161 POEPPDVTVLTSMLTDPSSHITAEAKRLARGSPPSLASSASQLSAPSKATCTTHD 2220
2221 SPDAOLIEANLLWRQEMGNITRVESENKVVILDSFEPDLHABGDERBISVAABILKRSK 2280
2221 SPDAOLIEANLLWRQEMGNITRVESENKVVILDSFEPDLHABGDERBISVAABILKRSK 2280
2281 FPSALPIWARPDPYNPLLESKWDPDVVPVHVHGCPLPPTKAPPIPPRRKRTVLTESNV 2340
2281 PAPALPWWARPDPYNPLLESKWDPDVVPVHVHGCPLPPTKAPPIPPRRKRTVLTESNV 2340
2341 SSALAEALATKTFSGSSSAVDSTATAPDLASDDGKSDVESYSSMPPLEGEFDPDL 2400
2341 PIALAEALATKTFSGSSSAVDSTATAPDLASDDGKSDVESYSSMPPLEGEFDPDL 2400
2401 SGGSNSTVGEBA SEDVCCSMYSYWTGALITPCHAABESKLPIINPLSNLLRHHNVAT 2459
2401 SGGSNSTVGEBA SEDVCCSMYSYWTGALITPCHAABESKLPIINPLSNLLRHHNVAT 2459
2460 TSSASLRQKKYTFEORLOVLDHYRDLKEMKAKASTVKAKLLSTEEACKLTPPHSAKSK 2519
2461 TSSASLRQKKYTFEORLOVLDHYRDLKEMKAKASTVKAKLLSTEEACKLTPPHSAKSK 2519
2520 FGYGAKDVENLSRRANVHRSVMEDELLEDTEFIDITTIMAKSEVFCVQPEKGRKPARLI 2579
2521 FGYGAKDVENLSRRANVHRSVMEDELLEDTEFIDITTIMAKSEVFCVQPEKGRKPARLI 2579
2580 VPFDLGVRVCEKALYDVVSTJFQAVNGSSYGFQVSPKORVEFLVNTWKSXKCPKGFSD 2639

Db 2581 VFDLGVRCVCKALYDVVSKLP:AYVWGSSYGQYSGQVRFVFLVQAWKSKTKPMGLSYD 2640
QY TRCFDSVTIVTESDIRVERSIYQCCDLAPBAEQATSLTERLIYICGPLTNSKGQNGCHRCR 2699
Db 2641 TRCFDSVTIVTESDIRVERSIYQCCDLAPBAEQATSLTERLIYICGPLTNSKGQNGCHRCR 2700
QY ASGVLTTCSCNTITCYLKATACRAAKLQDCTMLVNGDDLVWICESAGTQEDAAALRAFT 2759
Db 2701 ASRVLTTCSCNTITRYIKARAAACRAAGLQDCTMLVCGDDLWICESAGVQEDAAALRAFT 2760
QY EAMTRYGAPDPDPQPYDYLELTITSCSNVSAHADGKRVYLYLTPPTPLARAAWETA 2819
Db 2761 EAMTRYGAPDPDPQPYDYLELTITSCSNVSAHADGKRVYLYLTPPTPLARAAWETA 2820
QY RHTPINSWLGNIYKAPTLNARMTLMTHTPSSILLAOEKLKALDCQYGCYSIEPLDLP 2879
Db 2821 RHTPVNSWLGNIYKAPTLNARMTLMTHTPSSILLAOEKLKALDCQYGCYSIEPLDLP 2880
QY QIIEELHGLAFTLHVSYPGEINVASCLKLGVPPLRTWRHRARSVRKLLSOGGRAAT 2939
Db 2881 PIQRLHGLSNFSLHVSYPGEINVASCLKLGVPPLRTWRHRARSVRKLLSOGGRAAT 2940
QY CGRYLFNNAVRTKLPPIPAASQDLGSMFVAGYSGEDIYHLSLRARPRWFPCLLLLS 2999
Db 2941 CGRYLFNNAVRTKLPPIPAASQDLGSMFVAGYSGEDIYHLSLRARPRWFPCLLLLS 3000
QY 3000 VCVGYIYLPNR 3010
Db 3001 AGVGIYLLPNR 3011

RESULT 9
JC5620
Genome polyprotein - hepatitis C virus (isolate EUH1480)
N;Contains: capsid protein C; envelope protein E; hepatitis C virus (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: JC5620
R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M. Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant A;Reference number: JC5620; MUID:97366593; PMID:9223423
A;Accession: JC5620
A;Molecule type: mRNA
A;Residues: 1-3014 <C>A>
A;Cross-references: GI:Y13184
A;Experimental source: genotype 5a, which predominates in South Africa
A;Note: the translation of the nucleotide sequence is not complete in this paper
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <C>C>
F;116-191/Product: envelope protein E #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;384-408/Region: hypervariable #status predicted
F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F;1008-1616/Product: hepatitis C virus #status predicted <NS3>
F;1231-1238/Region: nucleotide-binding motif A (P-loop)
F;1313-1318/Region: nucleotide-binding motif B
F;1317-1320/Region: DEXH motif
F;1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F;1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F;2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 78.1%; Score 12496; DB 1; Length 3014;
Best Local Similarity 75.0%; Pred. No. 0;
Matches 2268; Conservative 309; Mismatches 423; Indels 24; Gaps 6;
QY 1 MSTNPKPORKTKRNTNRRPQDVKFGGQIVGGVYLLPRRPGKLVGRATRKNSRSQPRG 60

1 MSTNPKPORKTKRNTNRRPQDVKFGGQIVGGVYLLPRRPGKLVGRATRKNSRSQPRG 60
QY 61 RROPIKARPEBGRNAOQGYWPPLVXNGLWAGWLLSPGRSPWGPDPDRRRNRLG 120
Db 61 RROPIKARPEBGRNAOQGYWPPLVXNGLWAGWLLSPGRSPWGPDPDRRRNRLG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAAALAHGVREDDGVNYATGNLPGCSFSIFLIA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAFLGGAAALAHGVREDDGVNYATGNLPGCSFSIFLIA 180
QY 181 LLSCLIPASAYEVNRSVGIYHTNDCSSSIYVEADYIMHTPGCVPCVQEGNSRSCV 240
Db 181 LLSCLIPASAYEVNRSVGIYHTNDCSSSIYVEADYIMHTPGCVPCVQEGNSRSCV 240
QY 241 ALPTLALNARSVPPTTIRRHVDLLVGTAAFCASVYDGLCGSIFLVVSOLFTSPSRHET 300
Db 241 ALPTLALNARSVPPTTIRRHVDLLVGTAAFCASVYDGLCGSIFLVVSOLFTSPSRHET 300
QY 301 VQDNCISIYDGHVSGHRMAMMMNMSPTTALVVSOLLRIPOAVDMVAGAHGVLAGLA 360
Db 301 VQDNCISIYDGHVSGHRMAMMMNMSPTTALVVSOLLRIPOAVDMVAGAHGVLAGLA 360
QY 361 YYSVGNWAKVLIALLPAGVDSGTHTTGEVAGHTTSGPTSLFSSGASOKIQIVNTNGSW 420
Db 361 YFASTANWAKVLIALLPAGVDSGTHTTGEVAGHTTSGPTSLFSSGASOKIQIVNTNGSW 420
QY 421 HINRTALNCDLSQTGFALFYAHKFNSSGCGPERMASCRPIDWFAQGGMPITY-TKPN 479
Db 421 HINRTALNCDLSQTGFALFYAHKFNSSGCGPERMASCRPIDWFAQGGMPITY-TKPN 479
QY 480 SDQRPYCHYAPRPGVVPASQVCGPYCTPSPVVVGTDRSGVPTYSNGENETDMLL 539
Db 480 SDQRPYCHYAPRPGVVPASQVCGPYCTPSPVVVGTDRSGVPTYSNGENETDMLL 539
QY 540 NNTPPCCNWFCTWMSSTFTKCGPPNCGVGNRTLICTDCFRKHEPTATYKCGS 599
Db 541 NNTPPCCNWFCTWMSSTFTKCGPPNCGVGNRTLICTDCFRKHEPTATYKCGS 600
QY 600 GPWLTPCLDYVYRLWHYCTNLNFSIFKVMVVGVEHRLNAAKNWTRGERCNLEDRDR 659
Db 601 GPWLTPCLDYVYRLWHYCTNLNFSIFKVMVVGVEHRLNAAKNWTRGERCNLEDRDR 660
QY 660 SEISPLLSSTENQIPLCAFTTTPALSTGLIHLHONIVDQYLYGVGSAPVSAIKWEYI 719
Db 661 AELSPLLHTTQWAILPCSTPTPALSTGLIHLHONIVDQYLYGVGSAPVSAIKWEYI 720
QY 720 LLSFLILADARVACACLMMLLIQAALLENVYLANASVAGAHGILSLVFPFCAAWYK 779
Db 721 MLVFLILADARVACACLMMLLIQAALLENVYLANASVAGAHGILSLVFPFCAAWYK 780
QY 780 GRAPGAAYAFYGVWPFLLLLALPPRAYALDREMAASCGGAVLGLVFLTLSPYKVF 839
Db 781 GRAPGAAYAFYGVWPFLLLLALPPRAYALDREMAASCGGAVLGLVFLTLSPYKVF 840
QY 840 TRILWLOVYITRAAHMOWVPLNVGRDAILITLCAVHPELFDITKLLAILGLPL 899
Db 841 TRILWLOVYITRAAHMOWVPLNVGRDAILITLCAVHPELFDITKLLAILGLPL 900
QY 900 VYLOAGITRYPYFVRAQGLTRACMLVRKAVAGHVQVFMKLGALCTGVYVNHLPDRDW 959
Db 901 VYLOAGITRYPYFVRAQGLTRACMLVRKAVAGHVQVFMKLGALCTGVYVNHLPDRDW 960
QY 960 AHAGRLDRAVAVFPVPSAMETKVIWGTADTAACGDIILGLPVSAARRKEIFLGPADSLE 1019
Db 961 AHAGRLDRAVAVFPVPSAMETKVIWGTADTAACGDIILGLPVSAARRKEIFLGPADSLE 1020
QY 1020 GQGWELLAPITAYSQQTRGVLCIITSLTQDKNQVEGVQVSTAPQSFATCINGVW 1079
Db 1021 GQGWELLAPITAYSQQTRGVLCIITSLTQDKNQVEGVQVSTAPQSFATCINGVW 1080
QY 1080 TVYHGASKTLGPKGFIOTQYINVDLIVGQAPPGARSMITCSGSSSFLYVTRHADY 1139
Db 1081 TVYHGASKTLGPKGFIOTQYINVDLIVGQAPPGARSMITCSGSSSFLYVTRHADY 1140

QY	1140	IPVRRGDSRGSLSPRPVSILKSSGGPLLCPSGHVGVFRAAVCTRGVAKAVDPIVE	1199
Db	1141	IPARRGDTRASLLSPRPISYLKSSGGPIMCPSGHVGVFRAAVCTRGVAKALEZVPVE	1200
QY	1200	SMETTWESPVFTDNDSTPPAVPOTFOVAHLEFAPGSGKSTKVPAAVAQGVKVLVLPNSVA	1259
Db	1201	NLETTWESPVTNDSTPPAVPHEFQVGHLLHAPGSGKSKVPAAVAQGVKVLVLPNSVA	1260
QY	1260	ATLFGAYMSKAHGDINIRTVRTITGGSITYSTYKFLADGGSGGAYDIIICDECH	1319
Db	1261	ATFGGAYMSRAXGVDNIRTVRTITGAGITVSYTKFFADGGSGGAYDIIICDECH	1320
QY	1320	STDSITLIGTVLDQAEETAGRLVULATATPGSVTVPHPNLEETGLSNNGRIPSWGKA	1379
Db	1321	SQDATTILGTVLDQAEETAGRLVULATATPGSVTVPHPNLEEVNLPSEGBIPFYGRA	1380
QY	1380	IPEAIKGGRHILFCHSKKXCDLAAKILGLNLNAVAYVGLDVSIVPIPGDVVVWATDA	1439
Db	1381	IPLVLKGGRHILFANQKKKAKETAKMKNGPKRKAIVYVGLDVAVIPAIGDVVVWSTDA	1440
QY	1440	LMTGFTGDFDSVIDCNTCVTQTVDFSDPTPTTETTTVPDADVRSRORRGTOGRASGIY	1499
Db	1441	LMTGFTGDFDSVIDCNSAVTQTVDFSDPTFTTETTTVPDAYSRSORRGTOGRGRHIY	1500
QY	1500	RYFTGGRPSGDFSSVLCBCYDAGCAWYELTPAETSVRLRAYINTPGLPVCODHLEFWE	1559
Db	1501	RYVSSGERPSGIFDSVVLCEYDAGCAWYDLTPAETTVRLRAYINTPGLPVCQSHLEFWE	1560
QY	1560	SVFTGLTHIDAHFLSQTKQAGDNFPYLIVAYQATVCARAQAPPPSWDMWKCLIRLKPTLH	1619
Db	1561	GVFTGLTNIHAHMLSOAQKGSEFPYLIVAYQATVCVRAXAPPPSWDMWKMLCLKPTLT	1620
QY	1620	GPTELLYALGAVQNEVILTHPIKTYIMACMSADLEVUTSTWVLGGVLAALAAVCLTGTGS	1679
Db	1621	GPTELLYRLGAVQNEILTTHPIKTYIMACMSADLEVUTSTWVLGGVVAALAAVCLTVGS	1680
QY	1680	VWVGRIILSKPAVVPDREVLVYQEFDEMEECASQLEYIBQGMQLAEQFKQKALGLLOTA	1739
Db	1681	VAVVGRIILSGRPAITPDREVLVYQOFDEMEECSASLSEYVDEARAIAGQFKEKVLGLICTA	1740
QY	1740	TKQAEAAAPVVESKWRALEFTWAKHMNFISGIQYLACLSTLPCNPAIASIMAFASITS	1799
Db	1741	GQKAETILKPAATSMWSKABQWAKHMNFVSGIQYLAGESTLPCNPAVAILMSESTAATS	1800
QY	1800	PLTTQNTILLNILGCVWAAQALPPSAASAFVAGIAGAAVGSIGLGVKVLVDILLAGYAGV	1859
Db	1801	PLTTHTQILLNILGSWASQIAPTATAFVVSWMAGAVGNIGLGRVLIDILLAGYGTGV	1860
QY	1860	AGALVAFKWSGHEVPSTEDLWNLIPAILSPGALVGVVCAAILRRHVGPGEAVQWNNRL	1919
Db	1861	AGALVAFKMGESPTAEELWNLLPSILCPGALVGVVCAVLRHRTGPEGAVQWNNRL	1920
QY	1920	IAPASRGNHVSPTHVPSDAAAARTOILSLTITQLLKRLHOMINDEDCSPQSGMLRD	1979
Db	1921	IAPASRGNHGSPTHVPTDASAKVTQLLSLITVTSLLKRLHTWIGEDYSTPCDGTWLR	1980
QY	1980	VWDMICTVLTDKTVLQSKLPLRIPGVPPFTSCQSGYKGWVRGDGIMOTTPCGAQIAGHV	2039
Db	1981	IWDWVCTALTDFKAWLOAKLLPOLPGVPFFSCQSGYKGWVRGDGWNSTKPCPGATISGHV	2040
QY	2040	KNGSMRIVGPETCSNTHGTEPINAYTTGCTPSPADNYSALWRVAEEVVEVTRVGDV	2099
Db	2041	KNGTNRIIVGPLKCSNTHGTEPINATTTGSPVPAPPNYKALWRVGHADFAEVRVGDY	2100
QY	2100	HYVTGMTDNNKCPQCPVPAPEFFTEVDGRLHRVAPACKPLLRDVTVPQVLNLYVGSQ	2159
Db	2101	HYITGVQDNLKCPQCPVSPPEFFTEYLDGVRHRRPAPCPNPLLRBEVTFVSLHSYVVGSG	2160
QY	2160	LPCBPBPVTVLTMITDPSHIETATKRRLARSGSPPSLAAGSSASQLSAPLSKATCTTHH	2219
Db	2161	LPCBPBPVTVLTMISDPAHITETAKRRLNRRGSPPSLANSSASQLSAPLSKATCTIQG	2220

QY	2220	DSPDADLIEANLWRQCMGNNIIRVSEKNKVILDSPEPLAHAGDDEREISVAABAILKRSR	2227
Db	2221	HHPDADLIEANLWRQCMGNNIIRVSEKNKVILDSPEPLAHAGDDEREISVAABAILKRSR	2228
QY	2280	KFPFSAPIWARPDYNNPPLLESWKDPDYPPVPMVHAGCFJPTTKAPPIPPPRKRK-TWVLTES	2338
Db	2281	AFPPALFVMAEPGVYDPLELTKWKFDPYDPQVWGCPPIPPAGPDPVPLPRKRKPMELSDS	2340
QY	2339	NVSQALAEALATKTF-----GSSGSSAVDSGTATLPLDASDCDGKSDVESYS	2366
Db	2341	TVSQWADLADARFKVDPTPIEGQDSALGTSSSHDSG-----PEEKDD--NSDAASY	2392
QY	2387	SMPLEGECPGDPDLSGGSWSTVSEEAESVYVCSMSYTWTKGALITPCAAEESKLPINPLS	2446
Db	2393	SMPLEGECPGDPDLSGGSWSTVSGE--DNVVCMSYTWTKGALITPCSABEELPINPLS	2450
QY	2447	NSLLRHHNMVYATTSRSASLRQKVTFDRLQVLDDHYDRVLEKMKAKASTVKAKLLSIEE	2506
Db	2451	NLLRHHNLVYSTSRSGALRQKVTFDRLQVLDDHYDRVLEKMKALSKVKAKLLPLEE	2510
QY	2507	ACKUTPHPSAKSKPGYGAQOVNRLSSRAVNHIRSWEDELLEDTPDITTIMAKSEYPCV	2566
Db	2511	ACGUTPHPSARSKYGYAKVRSLLDKALKHIEGWQDLDDSDTPIPTTIMAKNEFVAF	2570
QY	2567	QPEKGGKPKARLIVPDLGVVCEKMALYDVVSTLPOAVMGSSYGFQYSPKQRFVFLVNT	2626
Db	2571	EPSKGGKPKARLIVPDLGVVCEKRALYDVAQKLPTALMGFSYGFQYSPAQRVDFLLKA	2630
QY	2627	WKSXKCPMGFSYDTRCFDSTVTESDIRVEESIYQCCDLAPPEAROAISLTLBELVIGGLPT	2688
DC	2631	WKSXKI PMAFSYDTRCFDSTITEHDIMTEESIYQCDLQPEARVAIRSLTORLYCGGPMY	2690
QY	2687	NSKQONCGYRRCRASGVLTTSCGNLTLCYLKATAACRAAKLODCTMLVNGDDLAVI CBSA	2746
Db	2691	NSKQOQCGYRRCRASGVFTSMGNMTFCYIKALASCAAKLUKDCITLLVCGDDLVAI CBSQ	2750
QY	2747	GTQDDAALRAFTEAMTRYSAAPPDGPQPEYDLELITSCSNVSVAHDDASGRVYVLTED	2806
Db	2751	GTHEDEASLRAFTEAMTRYSAAPPDGPVPVAYDLELITSCSNVSVAHDDASGRVYVLTED	2810
QY	2807	PTTPLARAWSTANUTPINSWLGNHIMVAPTLWARMILMTHFESSILLAOEOLKALDOQI	2866
Db	2811	PQVELAKAAWETAHSPVNSWLGNIIMVAPTLWARI VLMTHFVSLOSOGOLKTLAREM	2870
QY	2867	YGACYSIEPFLDLPOLIEBLRHLGSAFTLHSYSPGEINRVASCLRLKLGVPPLTWHRARSV	2922
Db	2871	YGSVSVTFPLDLPALIQRLHGLSFAFS_LHSYSPSEINRVASCLRLKLGVPPLRAWHRARAV	2930
QY	2927	RAKULLSQGGAATCGRY_FJNWAIVTKLKTPIPAASQDLDSGFVAGYSGGDIYHLSLRA	2988
Db	2931	RAKJLAQGGRAAICGJLYLFNWAIVTKRKLTPLDADADRLDLSWFTVGAGGGDIYHMSRA	2990
QY	2987	RPRWFPCILLLSVGVGHYILIPNR 3010	
Db	2991	RPRNLLLCILLLSVGVGHFLIPAR 3014	

RESULT 10

JQ1303
 Genome polypeptide - hepatitis C virus (isolate HC-J6)
 N1:Contains: capsid protein C; envelope protein M; hepatitis virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C1:Species: hepatitis C virus
 C2:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
 C3:Accession: JQ1303
 R1:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Izuka, H.; Machida, A.; Miyakawa, Y.,
 J. Gen. Virol. 72, 2697-2704, 1991
 A1:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
 A2:Reference number: JQ1303; MUID:92044440; PMID:1558196
 A3:Accession: JQ1303
 A4:Molecule type: Genomic RNA
 A5:Residues: 1-3033 <OK>
 A6:Cross-references: GB|D00944; NID:0221650; PIDN:BAAC00792.1; PID:G221651

A:Experimental source: isolate HC-76 from a Japanese individual		
C:Superfamily: hepatitis C virus genome polyprotein		
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polypeptide; serine proteinase; transmembrane		
F:2-115/Product: capsid protein C #status predicted <CPC>		
F:116-191/Product: envelope protein E #status predicted <EPM>		
F:192-389/Product: major envelope protein E #status predicted <MBE>		
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>		
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>		
F:1011-1619/Product: hepatitis C virus nonstructural protein NS3 #status predicted <NS3>		
F:1316-1321/Region: nucleotide-binding motif B		
F:1320-1323/Region: DEXH motif		
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>		
F:1867-2011/Product: nonstructural protein NS4b #status predicted <N4B>		
F:2018-3033/Product: nonstructural protein NS5 #status predicted <N05>		
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28		
Query Match	75.24; Score 12036.5; DB 1; Length 3033;	
Best Local Similarity	71.98; Pred. No. 0;	
Matches 2189; Conservative 336; Mismatches 472; Indels 45; Gaps 7;		
QY 1 MSTNPKPQRTKRNTRRRPQDKPFGGQIVGGVYLLPRGRPRGLVRAATKASRSQPRG 60		
DB 1 MSTNPKPQRTKRNTRRRPQDKPFGGQIVGGVYLLPRGRPRGLVRAATKASRSQPRG 60		
QY 61 RQDIPKARPEGKAWAQPGYFWPLNGEGLGWAGWLLSPRGSPSWGPTDPRRSNVLG 120		
DB 61 RQDIPKARPEGKAWAQPGYFWPLNGEGLGWAGWLLSPRGSPSWGPTDPRRSNVLG 120		
QY 121 KVIDTLTCGPFADLMGYPLVCAPIGGARALAHGVRLDGVNATGNLPCSCSIFLLA 180		
DB 121 KVIDTLTCGPFADLMGYPLVCAPIGGARALAHGVRLDGVNATGNLPCSCSIFLLA 180		
QY 181 LLSCLTIPASAEYERNVSGIHYHTNDCSNSIIVYEAADVMHTPCVPCVQEGNSSRCW 240		
DB 181 LLSCLTIPASAEYERNVSGIHYHTNDCSNSIIVYEAADVMHTPCVPCVQEGNSSRCW 240		
QY 241 ALPTLAAASVPTTIRRHVDLLVGTAPASAMVVDLCCGSLFVLSQLETFPRRHET 300		
DB 241 PVSENVAVQQGALQTGLRTHIDVMSATLCSALYVGLDCCGVMALAAQMFIVSPQHWF 300		
QY 301 VQDCNCSITPGHVSCHRMAMDMNWSPTTALVVSQILRIPOAVVDMVAGAHGWGLA 360		
DB 301 VQDCNCSITPGHVSCHRMAMDMNWSPTTALVVSQILRIPOAVVDMVAGAHGWGLA 360		
QY 361 YSMVGNWAKVLIIVALLFAGVDGTHTRVAGHTSGTSLFSSGASQKIQLYNTNGSW 420		
DB 361 YFSMQGAWAKVVILLAAAGVDAQTHTVGGSTAEANLTCMFLGARQKIQILNTNGSW 420		
QY 421 HINRTALNCNDSLOTGFAALFYAHKNSSCCPERMASCRPIDWFAQGWGPIITY--TKPN 478		
DB 421 HINRTALNCNDSLOTGFAALFYAHKNSSCCPERMASCRPIDWFAQGWGPIITY--TKPN 478		
QY 479 SSDORPCWHYAPRPGVVPASVCGPVYCTFPVGVVGTDRSGVPTYSNGENETDML 538		
DB 481 PEDKRPYCHWYPPQCGVVSASSVCGPVYCTFPVGVVGTDRSGVPTYSNGENETDML 540		
QY 539 LNNTRPQGNWFGCTWMNSTGFTTKCGPPCNIGGVGNRT--LICPTDCKRKHPEATYTK 596		
DB 541 LNSRTPQGSFEGCTWMNSTGFTTKCGPPCNIGGVGNRT--LICPTDCKRKHPEATYTK 600		
QY 597 CGSGPWLTPRCVDPYELHYPCTLNFSPKVMYVGVVHRLNAACNWTGRCNLED 656		
DB 601 CGSGPWLTPRCVDPYELHYPCTLNFSPKVMYVGVVHRLNAACNWTGRCNLED 660		
QY 657 RDRGELSPALLSTTEWQILPCAFVTLPALSTGLIHQNIIVDVOLYGVGSFAFVFAIK 716		
DB 661 RDRGELSPALLSTTEWQILPCAFVTLPALSTGLIHQNIIVDVOLYGVGSFAFVFAIK 720		
QY 717 EYIILLFLLADARVACLMMWLLAQAEALENLVNLNAASVAGAHGILSFVFFCAAW 776		
DB 721 EYIILLFLLADARVACLMMWLLAQAEALENLVNLNAASVAGAHGILSFVFFCAAW 780		
QY 777 YIKGLADGAAVAFVGVWPLILLLLALPPRAVALDREMAASCGGAVLVGLVFLTSLSPYIK 836		

781	YIKGRVPLATYSLTGLWSFGLLLALPOQAVAYDASVHGQICGAALLVLTITLTPFGYK 840	Db
837	VETRIIWMLCYFETRAEAHMVQVWVPLAVRGGGRDALIILTCVAVHPELFIIDIKLLAIL 896	QY
841	TLISRLFWLWLYLTTLAEAMVQWAPPMOVRGGDLIIWAVAFPCGVVDFDIKWLLAVL 900	Db
897	GPLMVLQAGITRYPYFVRAQGLIRACMLVKVAGGHYVQVYFKLGAITGVYVNHITPL 956	QY
901	GPAYLLKGLATRVYFVRAHALLRMCTWVHLAGGRVQVWLLALGRWTCTVIYDHLTPM 960	Db
957	RDWAHGLRLAVAPVVPVSAMETKVIITWCAATAAGDIIILGIPVARRGKELFPGPAD 1016	QY
961	SDWAANGRLRLAVAPVVPVSAMETKVIITWCAATAAGDIIILGIPVARRGKELFPGPAD 1020	Db
1017	SLRGQWMLAPITAYSQQRGVGLGIIITSLTGRDKXNVEGEVQVYVSTATQSFELATCING 1076	QY
1021	GYTSKGSLLAPITAYAQTRGLLGIITVMSVGRDQTEQAGEIQVLTSTVQSFELATCING 1080	Db
1077	VCWTVYHAGSKTLAGPKGSIITQMYTNVDLVLVGMQAPPGARMTPCSGSSDLVLTNRH 1136	QY
1081	VLMTVYHAGSKTLAGPKGSIITQMYTNVDLVLVGMQAPPGARMTPCSGSSDLVLTNRH 1140	Db
1137	ADVIPVRRRSGSLLSPVSYLXSSGGPILCPSGHVGVYFRAAVCTRGVAKAVDFI 1196	QY
1141	ADVIPARRRSGSLLSPVSYLXSSGGPILCPSGHVGVYFRAAVCTRGVAKAVDFI 1200	Db
1197	PVSEMTTSPVFTUNSTPPAVPQTFQVAHLHAPTSKSGKSTKVPAAAYAAQGVKVLNLP 1256	QY
1201	PVETLIVTESPTFSDNSTPPAVPQTFQVGYLHAPTSKSGKSTKVPAAAYAAQGVKVLNLP 1260	Db
1257	SVAATLFGAYMSKAHGDINITGVRTITGTSITVSYTKFLADGGSGGAYDIIICD 1316	QY
1261	SVAATLFGAYMSKAHGDINITGVRTITGTSITVSYTKFLADGGSGGAYDIIICD 1320	Db
1317	ECHSTDSITLIGTGLVLDQAEAGALVLTATPPGSVTPVPHNTEETGLSNNGIIPFY 1376	QY
1321	ECHAVDSTLIGTGLVLDQAEAGALVLTATPPGSVTPVPHNTEETGLSNNGIIPFY 1380	Db
1377	GKAIPTEAIKGGREHLIFCHSKKCKDELAALRGWGLNNAVAYRGLDVSIPIPGDVVVA 1436	QY
1381	GRAIPLSYIKGGREHLIFCHSKKCKDELAALRGWGLNNAVAYRGLDVSIPIPGDVVVA 1440	Db
1437	TDALMTGFTGDFPSVIDCNVAVTQVDFSLDPTFTTTTTPQDASVRSORGRGTGRS 1496	QY
1441	TDALMTGFTGDFPSVIDCNVAVTQVDFSLDPTFTTTTTPQDASVRSORGRGTGRS 1500	Db
1497	GIYRFTVPGRRPSPGSDSSVLCBCYDAGCAWYELTPAETSVRLRAYLNTPLPVCODHLE 1556	QY
1501	GIYRFTVPGRRPSPGSDSSVLCBCYDAGCAWYELTPAETSVRLRAYLNTPLPVCODHLE 1560	Db
1557	FWBSVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCAZAAQAPPPSWDMWKCLRLKP 1616	QY
1561	FWBAVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCAZAAQAPPPSWDMWKCLRLKP 1620	Db
1617	TLHGPTPLLRGLAVQNEVILTHPIIKYINACMSADLEWVTSVWLVGGVLAALAAAYCLT 1676	QY
1621	TLVGPTPLLRGLAVQNEVILTHPIIKYINACMSADLEWVTSVWLVGGVLAALAAAYCLT 1680	Db
1677	TGSVVIIVGRILISGKPAWPDREVLVQEPDEMEECASOLPIEQGMOLABQFQKALGLL 1736	QY
1681	TGCVCIIGRHVQRAVAVDPKDELVEAPDEMEECASRAALIIEGQRIABMLSKIQGLL 1740	Db
1737	QATKQAEAAAPVVESEKRALETFWAKHMNFISGTYLAGSLTLEPNPAIASLMAPTAS 1796	QY
1741	QAKSKAQDTPQAVQASQWPKVQWAKHMNFISGTYLAGSLTLEPNPAIASLMAPTAS 1800	Db
1797	ITSPLTQNTLLFNILGWWAAQALAPESASAFVAGIAGAAVGSIGLKVLDILAGY 1856	QY
1801	LTSPLSTTILLNIGWLLASQIADPAGATGFVVSGLVGAAYVGSIGLKVLDILAGY 1860	Db
1857	AGVAGALAVAKVSGEVPSTEDLVNLLPALISPCALVGVVCAILRRHVPGEAGVOWM 1916	QY

```
Db 1861 AGISGALVAFKIMSGEKPSMEDVWNLPGILSPGALVWGVCAAILRRHVGPGEAGVQM 1920
QY 1917 NPLIAPASGNHVSPTHYYPEDGAARVQIILSSLTITOLLKRLQWINECDSTPCSGSW 1976
Db 1921 NPLIAPASGNHVSPTHYYPEDGASQVQLLGSITITSLRLNHNITTEDCPIPCSGSW 1980
QY 1977 LRDVWDNICTVLTDFTWTQSKLLPRLPGVPPFLSCORGKGVWYRGDGIQWTCPCGAQIA 2036
Db 1981 LRDVWDVCTILTDFKNWLTSLKFLFKMPGLPFISCKGKGVWAGTGMITPCPGANIS 2040
QY 2037 GHVKGSMRIVGPRTCSTWHGTFPINA YTGCTPSAPNYSRALMRVAABEYVEVTV 2096
Db 2041 GNVGLGSMRITGKTCMNIWQGFPIPCYTEGQCVKPAFPKFIATRWAAAEYAEVTOH 2100
QY 2097 GDFHYVTGTTNWKCPQVAPERFTEVDGVRLHYAPACKPLLEDYTFQVGLNOXILV 2156
Db 2101 GSHYITGSLTDNLKVPCLPSPERFSWVGVOIHREFAPIPKPFFRDEVSFCVGLNSFVV 2160
QY 2157 GSQLPCEPDPDVTLTSMULTDPSHITAEAKRLARGSPPSLASSASQLSAPSLKATCT 2216
Db 2161 GSQLPCEPDPDVTLTSMULTDPSHITAEAKRLARGSPPSLASSASQLSAPSLKATCT 2220
QY 2217 THDSADLIEANLWRQBMGNITRVSEBKVILDSFEPHARGDERBISVAEILR 2276
Db 2221 THGKAYDVDMVDANLF---MGDVTRIESGSKVVLDSLDPMVEERSDLEPISPEYML 2276
QY 2277 KSKKFPSPALPIWAPDYNPELLESWDPDVPVWVHGCPLPPTKAPPIPPPRKRTWILT 2336
Db 2277 PKKRFPPALPAWAPDYNPELLESWKPDIQDPATVAGCALPPPKKTPPPRRRRVGLS 2336
QY 2337 ESNVSALAEALAKTRFG-----SSGSSAVDSGTATALPDLASDDGDKGSDVESY 2386
Db 2337 ESSIAALQALAKISFGQPPSGDGLSTGADAADSGSRTPPELAL-----SETGSI 2390
QY 2387 SMPPLSGEPDPI-----SGSKSTVSEBESADVVCMSMYTWT 2426
Db 2391 SMPPLSGEPDPILEPEQVLEPPQGVVTPGSGSGSWSTCSEE-DDSVCCMSMYSWT 2449
QY 2427 GALITPCAAESKLPINPLNSLIRHNHMYATTSASLRQKKVTFDRQLVLDHVRDV 2486
Db 2450 GALITCSPEEKLPINPLNSLIRHNHMYATTSASLRQKKVTFDRQALDAHVDSV 2509
QY 2487 LKEMKAASTVAKLJIEACKLTPPHSAKSPGYGAKDVRLSSAVNHISWEDLL 2546
Db 2510 LKDIKAASKVTA RLTLLEACQLTPPHSARSKYFGAKEYRSLSGRANHIKSWKDLL 2569
QY 2547 EDTETPTDTIMAKSEVFCVQPKGKPARLIVPDLGVVCKEMALYDVVSTLPOAVM 2606
Db 2570 EDTQTPPTTIMAKNEVFCVDPTKGGKKAARLIVPDLGVVCKEMALYDITQKLPOAVM 2629
QY 2607 GSSYGFQYSPKQVFEFLVNTWKSXKCPMGFSYDTRCFDSTVTSDIRVEESIYQCCLAP 2666
Db 2630 GASVGFQYSPAQRVEFLKAWAEKDPMGFSYDTRCFDSTVTSDIRVEESIYACSLPE 2689
QY 2667 EARAQIASLTERLVIGGLPLNSKQNGCYRRCRASGVLITSCGNLTLCYKATAACRAAK 2726
Db 2690 EAHTAIHSLTERLVIGGLPLNSKQNGCYRRCRASGVLITSMGNTITCYWKALAAACRAA 2749
QY 2727 LQDCTMLVNGDLVVICESACTODAAALRAFTAMTRYSAAPCPDPPQPEYDLELITSCS 2786
Db 2750 IIAFTMLVCGDDLVISESQTEDEENLRAFTAMTRYSAAPCPDPPQPEYDLELITSCS 2809
QY 2787 SNVSVAHDASGRVYVITRDTPTTLARAAMETARTHTPINSWLGNIIMVAPTLMARMLMT 2846
Db 2810 SNVSVALGPQGRYRYLTRDTPTTLARAAMETARTHTPINSWLGNIIMVAPTLMARMLMT 2869
QY 2847 HFFSIIAQEOLEKALDQIYGACVSTBPLDLPOILIERLHGLSAPTHLSYSPGINRVAS 2906
Db 2873 HFFSIIQAQDTLDQNLNPEMYGAVYSVSPDLPAIIERLHGLDAFSLHTYTPHLTRVAS 2929
QY 2907 CLRLGVPLPLTWRRHARSVRKLLSQGGRATCGRYLFWNAVTKLKTPIPAASQDL 2966
Db 2930 ALRLGAPPLKAWKRAKARVASLISRGRAAVCGRYLFWNAVTKLKTLPPEARLLDL 2989
```

2967 SGWFFVAGSGGDIYHSLSRARPRWFPECLLLLSVGVGIYLLPNR 3010

2990 SSWFFVAGSGGDIYHSLSRARPRWFPECLLLLSVGVGIYLLPNR 3033

RESULT 11

GNWVJ8

genome polyprotein - hepatitis C virus (strain HC-J8)

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001

C:Accession: A40250; PQ0397; PQ0559

R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.; Virology 188, 331-341, 1992

A:Title: Pull-length sequence of a hepatitis C virus genome having poor homology to repository reference number: A40250; MUID:92230232; PMID:1314459

A:Accession: A40250

A:Molecule type: genomic RNA

A:Residues: 1-3033 <OKA>

A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Pollett, E.; Yap, P.L. J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to repository reference number: PQ0393; MUID:92268871; PMID:1316939

A:Accession: PQ0397

A:Molecule type: genomic RNA

A:Residues: 2678-2754 <CHA>

A:Cross-references: DBJ:D10134

A:Experimental source: Isolate B-12

R:Kato, N.; Ootsuyama, Y.; Okoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno, Biochem. Biophys. Res. Commun. 181, 273-285, 1991

A:Title: Distribution of plural HCV types in Japan.

A:Reference number: PQ0554; MUID:92068204; PMID:1720309

A:Accession: PQ0559

A:Molecule type: mRNA

A:Residues: 2678-2729 <KAT>

A:Cross-references: GB:D10562; GB:D30518; NID:g221523; PIDN:BAA01418.1; PID:g221524

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein; capsid protein C #status predicted <CPC>

F:115/Product: capsid protein M #status predicted <BPM>

F:116-191/Product: major envelope protein E #status predicted <MEB>

F:192-389/Product: nonstructural protein NS1 #status predicted <NS1>

F:390-733/Product: nonstructural protein NS2 #status predicted <NS2>

F:734-1010/Product: nonstructural protein NS3 #status predicted <NS3>

F:1011-1619/Product: hepacivirin #status predicted <NS3>

F:1234-1241/Region: nucleotide-binding motif A (P-loop)

F:1316-1321/Region: nucleotide-binding motif B

F:1320-1323/Region: DEXH motif

F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>

F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>

F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,233

Query Match 74.8%; Score 11972.5; DB 1; Length 3033;

Best Local Similarity 70.9%; Pred. NC. C;

Matches 2158; Conservative 369; Mismatches 473; Indels 43; Gaps 9;

QY 1 MSTNPKPQKTKNTNRFPQDVKFGGQIVGGVLLPRRGLGVRAIKASERSQPRG 60

Db 1 MSTNPKPQKTKNTNRFPQDVKFGGQIVGGVLLPRRGLGVRAIKASERSQPRG 60

QY 61 RRQFIPKARPEGRAWAQFGYPMPLVYGNBGLWAGWLLSPRSGRPSWGTPDPRRRNLG 120

Db 61 RRQFIPKARPEGRAWAQFGYPMPLVYGNBGLWAGWLLSPRSGRPSWGTPDPRRRNLG 120

QY 121 KVDTTCGADLMGIVPLVGAIPGLGAARALAHGVLEEDVNNVATGNLPGCSFSIFLLA 180

Db 121 RVDTTCGADLMGIVPVVGAIVGGVARALAHGVLEEDVNNVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTIPASAYEVNVSIGYHYVTNDCSSNSIIVVEAADVTMTGCVPCVQEGNSRRCW 240

Db 181 LLSCVTVPVSAVEVRNISSYYATNDCSNNSITWQITDAVLHLPGCVPCENDNGNLHCWI 240
 QY 241 ALTPPLAARNASVPTTIRRHVLDLVGTAAAFCSAMYVGDLCGSIPLVSQLFTFSRREHET 300
 Db 241 QVTENVAVKRGALTRSLRTHVDMITVMAATACALYVGVGCGAVMLQAQFVWSQRNF 300
 QY 301 VQDCNCSYIPGHVSHRWADNMGWSTTALVVSQILIRIQAVVDMVAGAHGWGLAGLA 360
 Db 301 TQECNCSYIQGHITGHRWADNMLWSSEPTLIMLAYAARVPELVLEIIFGGHGWVVFGLA 360
 QY 361 YYSWGNWAKVILVALLFAGVDGETHTTGRVAGHTTSGFTLSFGSASQKIQLVNTNGSW 420
 Db 361 YFSMQAWAKVIAILLVAGVDATYSSQAGRIVAGFAGFYTGAKQNYLYLNTNGSW 420
 QY 421 HINTALNCNDSLOTFGPAALFYAHKFNSSGCPERMASCRPIDWFAQGWGPITY--TKPN 478
 Db 421 HINTALNCNDSLOTFGLASIFYTHKFNSSGCPERLSSCRGLDDFRIGWGLTEYETNVN 480
 QY 479 SSDRPIYCHWYAPPCGVVPASQVGYCFTPTSPVVGTTDRSGVYTSYSGENETDVML 538
 Db 481 DGDWRPIYCHWYPPRPGIIVPARTVCGPIYCFPTSPVVVGTTDKQGVPTTWGENETDVF 540
 QY 539 LNNTRPQGNWFGCTWNNSTFTKTCGGPPCNIIGGVGNRT--LTCPTDCRKHPEATYTK 596
 Db 541 LNSTPRPRGAWFGCTWNGTGFTKTCGAPPCRIRKDYNSITDILLCPTDCRKHPEATYTK 600
 QY 597 CGSGPWLTPRCLVDYPYRLWHYPCITNFSIFKVMYVGGVEHRLNAAACNWRGRNCLED 656
 Db 601 CGAGPWLTPRCLVDYPYRLWHYPCVNTFTIFKARMYVGGVEHRSFSAACNTRGRDCSLED 660
 QY 657 RDRSELPLLLSTTEWQILPCAFITLTPALSTGLHLHQNIVDQYLYGLSPALTRYIVKW 720
 Db 661 RDRGQSPULLHSTTEWAVLFCSDDLTPALSTGLHLHQNIVDQYLYGLSPALTRYIVKW 720
 QY 717 EYLLLEPLLADARVACLMWMLLIAQEAALENLVYNAASVAGAGHILSFLVFFCAAM 776
 Db 721 EWVILLFLLADARICACLMWMLLIAQEAALENLVYNAASVAGAGHILSFLVFFCAAM 776
 QY 777 YIKGRILPAGAAFYGVWPIILILLALIPPRAYALDREMAASCGAVLVGLVITLSFYXK 836
 Db 781 YIKGRVVPVATSVGLWGLFLLVIALPOQAVALDAEQEGLALIVITISITFLTPAYK 840
 QY 837 VFLTRILMWLOVPIITRAEAHMOWVPELVNVRGGRDAIILLTCVHPELIFDITKLLAIL 896
 Db 841 ILLSRVWMLSYMVLAEAQIQWVPLEVRGGRDGIWVAILHPLRVFVTKWLLAIL 900
 QY 897 GLMWLOAGITVPVVRACGLIRACMLVRKAGHYVQVMFMKLGALGTGYVWHLTPL 956
 Db 901 GPAYLLKASLIRIPVRAHALLRVCTLVKHLAGARYIQWLLITIGRWGTGYIYDHLSP 960
 QY 957 RDWAHAGRLDAVAVBPVVSAMETKVIITWADTAACGDIILLGLPVSARRKEIFLGPAD 1016
 Db 961 STWAAGLRLDLAIAVEPVVFSNKKVIVMGARTVACGDIHLGLPVSARLGRVLLGPAD 1020
 QY 1017 SLEGGOWRLLAPITAYSQOTRGVLGCIITSLGRDNKOVEGVVSTATQSFATCING 1076
 Db 1021 GYTSXGKLLAITAYTQOTRGLLGAIVVSLGRDNKAGQGVQVLSSTVQTFGLTSISG 1080
 QY 1077 VQWTVYHAGSKTLAGPKPTQMTVNTVDLIVGWAOPPGARSMTPCSGSSDLIVYWRH 1136
 Db 1081 VLWTVYHAGNKTLAGPKGPVQMTVSAEGDLVGNFSPGPKSLDPCGAVDLVYTRN 1140
 QY 1137 ADVIPVRRGDSRGSLLSPRVSVLYKSGSGPILCPSHVGVVFRANVCTRGVAKAVDFI 1196
 Db 1141 ADVIPVRAKDRRGALLSPRJLSTLKGSGGVPVLCRSHAVGLFPAAYCAARGVAKSIDFI 1200
 QY 1197 PVBSNETTMRSPVFTDNTPTPAVQPTFOVAHLHAPTGSKSTKVPAAVAAQYKVLVLP 1256
 Db 1201 PVESLDVATRPSPSDNSTPTPAVQSYQVYLHAPTGSKSTKVPAAVAAQYKVLVLP 1260
 QY 1257 SVAATLFGAYMSKAHGDIPNIRTVRTITGGSTYTYGKFLADGCGSGAYDIILICD 1316
 Db 1261 SVAATLFGAYMSKAHGINPRTVTVTTGDSITYTYGKFLADGCGACAGAYDIILICD 1320

QY 1317 BCHSTDSTTILIGITVLDQAGTARLVLATATPPGCVTVPHNIBIIGSNNGEIPFY 1376
 Db 1321 BCHSVDTATILIGITVLDQAGTARLVLATATPPGCVTVPHNIBIIGSNNGEIPFY 1380
 QY 1377 GKAIPIEAIKGRHLIFCHSKKKCDELAALKITGLGNAAVYRGDLSVPIPIGDVVVA 1436
 Db 1381 GKAIPLAFI KGRHLIFCHSKKKCDELAALRGVNAVYRGDLSVPIPTQGDVVVVA 1440
 QY 1437 TDALMTGTFGDFSDVIDNCITCVTVDPFSLDPTFTTITTTVPQDAVSRRGRGTGRS 1496
 Db 1441 TDALMTGTFGDFSDVIDNCVAVSQIVDFSLDPTFTTITTTVPQDAVSRRGRGTGRS 1500
 QY 1497 GIVRFVTPGRRPSCMFDSVLCBYDAGCAWYELTTRATSVRLRAYLNTTGLPVCQDHLE 1556
 Db 1501 GVTRYSSGRRPSCMFDSVLCBYDAGCAWYELTTRATSVRLRAYLNTTGLPVCQDHLE 1560
 QY 1557 FWESVFTGLTHIDAHFLSOTKQAGDNPPYLVAYOATVCARAQAPPPSWDMWKLRLKP 1616
 Db 1561 FWEAVFTGLTHIDAHFLSOTKQAGDNPPYLVAYOATVCARAQAPPPSWDMWKLRLKP 1620
 QY 1617 TLHGPTLLYRLGAVONEVILLTHPIKYIMACMSADLEVVTWTVLGGVLAALAAAYCIT 1676
 Db 1621 TLHGPTLLYRLGAVTNEVTLTHPIKYIATCMQADLEINTSSWLAGGVLAALAAAYCIT 1680
 QY 1677 TGSWVIWGRILSGKPAVVPDREVLVYOEDEMEBECASQVPIBQGMQLABQFKOKALGLL 1736
 Db 1681 TGCISIIIGRHLNDRVVVAPDREVLVYOEDEMEBECASQVPIBQGMQLABQFKOKALGLL 1740
 QY 1737 QATKQAEAAAPVYESKWRALFTFWAKHWNFTISGIQYLAGLSTLPGNPAISLMAFTAS 1796
 Db 1741 QOATROAQDIOPAIQSSWPKLFQFWAKHWNFTISGIQYLAGLSTLPGNPAISLMAFTAS 1800
 QY 1797 ITSPLTQNTLLENLGGWVAQLAPPSSAASAPVAGIAGAAVSGISGLKVLVDILAGYG 1856
 Db 1801 LITSPLTSTILLNMGWLASQIAPPAGATGVVSGLVGAAGVSGISGLKVLVDILAGYG 1860
 QY 1857 AGVAGALVAFKWSGEPSTEDVNLIPAILSPGALVGVVCAAILRRHVGGEGAVQVM 1916
 Db 1861 AGLSALVAFKWSGEPSTEDVNLIPAILSPGALVGVVCAAILRRHVGGEGAVQVM 1920
 QY 1917 NRIIAFASRGNHVSPTHYVPESDAAARVTOILSSITITOLLKRLHOWNEDCSTPCSGSW 1976
 Db 1921 NRIIAFASRGNHVAETHYVPESDAASQRTQVILSSITITSLRRLHANITEDCPCSGSW 1980
 QY 1977 LEDVWDNTCTVLTDEKTMQSKLLPRLPGVPLSCQRYGKGVWRGDGIMQITPCGAQIA 2036
 Db 1981 LQDIWDMVCSILITDFKNWLSKLLPKMFGIPFISQCKYKGVWAGTGVWTTTRCPOGANIS 2040
 QY 2037 GHVKNMGSRIVGPRTCSTNTHGTPEINATTTGCTPSPAPNYSRALMRVAAEYVEVTRV 2096
 Db 2041 GHVRMTWTKITGPKICLNLWOGTTPINCITGCPVPKPPPNYKTAIRKVAASEYVEVTQH 2100
 QY 2097 GFHVYVTGMTDNNKCPQCPVPAFPFTTVDQVRLHRVAPACKPLLRDVTPOVGLNOVLV 2156
 Db 2101 CSFSVYVTLGTSNLLKVPQVPAPEFFSWVDVGVQIHRFAPVGPFPFRBEVFTVGLNFPV 2160
 QY 2157 CSQPCBCEPPEPVTVLTSMITDPSHITBETAKRRLARGSPPLSASSASQISAPSLKATCT 2216
 Db 2161 GSQPCPCEPPEPTEVLASMLTDPSHITAEAAARLARGSPPLSASSASQISAPSLKATCT 2220
 QY 2217 THHDSPADLLEALLWROEMGNI TRVESNKVILDSFELHARGDEHISVAABELR 2276
 Db 2221 THKAYDCDVMANLF---MGGDVTFIESDSKVIILDSLOSMTTEVEDDEPSPSYLI 2276
 QY 2277 KSRKFPSPALPIWAPDYNPPLLESKDPDYVVPVWHGCPPLPPTKAPPPIPPRRKKTVLT 2336
 Db 2277 KRRKFPSPALPEWAPDYNPVLITWKPVGPEPTVILGALPPTPOTPVPPRRRAKVL 2336
 QY 2337 ESNVSSALALATKTFG-----SGSSAVDSGTATLALPOLASDDGDK--GSDVESYSS 2387
 Db 2337 QDNVEGLREMAKVLSPLODNNDSGHSTGADTG---GDIVQOPSDETAASEAGSLSS 2391

QY	1	MS	NP	KP	QK	KT	NT	NR	PD	VK	FG	GG	QI	VG	SV	LL	PR	GP	PL	GR	PR	GL	VR	AT	KAS	ER	SQ	PR	G	60																									
Db	1	MS	NP	KP	QK	KT	NT	NR	PD	VK	FG	GG	QI	VG	SV	LL	PR	GP	PL	GR	PR	GL	VR	AT	KT	SER	SQ	PR	G	60																									
QY	61	RR	Q	P	I	K	AR	R	E	G	RA	W	A	Q	P	C	Y	P	W	P	L	Y	G	N	E	G	L	G	W	A	120																								
Db	61	RR	Q	P	I	K	AR	R	E	G	RA	W	A	Q	P	C	Y	P	W	P	L	Y	G	N	E	G	L	G	W	A	120																								
QY	121	KV	ID	LT	TC	GF	AD	LM	G	Y	I	P	U	G	A	P	L	G	A	R	A	L	A	H	G	V	R	V	E	D	G	180																							
Db	121	KV	ID	LT	TC	GF	AD	LM	G	Y	I	P	U	G	A	P	L	G	A	R	A	L	A	H	G	V	R	V	E	D	G	180																							
QY	181	LL	SC	LT	II	P	A	S	A	E	V	R	N	S	G	I	Y	H	V	T	N	D	C	N	S	I	V	E	A	A	D	V	IM	HT	PG	C	V	Q	E	G	N	S	R	C	W	240									
Db	181	LL	SC	LT	II	P	A	S	A	E	V	R	N	S	G	I	Y	H	V	T	N	D	C	N	S	I	V	E	A	A	D	V	IM	HT	PG	C	V	Q	E	G	N	S	R	C	W	240									
QY	241	AL	T	P	T	L	A	R	N	S	V	P	T	T	I	R	R	H	V	D	L	L	V	G	T	A	A	F	C	S	A	M	Y	V	G	D	L	C	G	S	I	F	V	S	O	L	F	T	P	S	R	H	E	T	300
Db	241	AL	T	P	T	L	A	R	N	S	V	P	T	T	I	R	R	H	V	D	L	L	V	G	T	A	A	F	C	S	A	M	Y	V	G	D	L	C	G	S	I	F	V	S	O	L	F	T	P	S	R	H	E	T	300
QY	301	VD	C	N	S	I	P	G	H	V	G	H	R	W	A	D	N	M	N	S	P	T	A	L	V	S	Q	L	L	R	I	P	G	A	V	D	V	M	A	G	A	H	G	V	L	A	G	A	360						
Db	301	VD	C	N	S	I	P	G	H	V	G	H	R	W	A	D	N	M	N	S	P	T	A	L	V	S	Q	L	L	R	I	P	G	A	V	D	V	M	A	G	A	H	G	V	L	A	G	A	360						
QY	361	Y	S	M	V	G	N	A	K	V	L	I	V	A	L	F	A	G	V	D	G	E	T	T	G	R	V	A	G	H	T	T	S	G	F	S	I	F	S	G	A	S	O	K	I	Q	V	N	T	N	G	S	W	420	
Db	361	Y	S	M	V	G	N	A	K	V	L	I	V	A	L	F	A	G	V	D	G	E	T	T	G	R	V	A	G	H	T	T	S	G	F	S	I	F	S	G	A	S	O	K	I	Q	V	N	T	N	G	S	W	420	
QY	421	H	N	R	T	A	L	N	D	S	L	O	G	F	P	A	A	L	F	A	H	K	F	N	S	G	C	P	E	R	M	A	S	C	R	P	I	D	M	F	A	Q	M	G	P	I	T	Y	T	K	P	N	S	480	
Db	421	H	N	R	T	A	L	N	D	S	L	O	G	F	P	A	A	L	F	A	H	K	F	N	S	G	C	P	E	R	M	A	S	C	R	P	I	D	M	F	A	Q	M	G	P	I	T	Y	T	K	P	N	S	480	
QY	481	D	R	P	C	M	H	A	D	R	P	C	G	V	P	A	S	O	V	C	G	P	V	Y	C	F	T	P	S																										

A;Experimental source: isolate JK5
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F;1-191/Product: core protein #status predicted <MAT1>
 F;192-383/Product: envelope protein 1 #status predicted <MAT2>
 F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 24.7%; Score 3952; DB 2; Length 782;
 Best Local Similarity 90.9%; Pred. No. 1.4e-232;
 Matches 711; Conservative 31; Mismatches 40; Indels 0; Gaps 0;

QY 1 MSTNPKPQRTKNTNRRPQDVKPGGQIVGGVYLLPRRGPRLGVRATRKASERSQPRG 60
 DB 1 MSTNPKPQRTKNTNRRPQDVKPGGQIVGGVYLLPRRGPRLGVRATRKATRSERSQPRG 60

QY 61 RROPPIKARRPEGRAWAQPQGYPMPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRSRNLG 120
 DB 61 RROPPIKARQPEGRAWAQPQGYPMPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATNGLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADPMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATNGLPGCSFSIFLLA 180

QY 181 LLSCLTTPASAYEVNRVSGIYHVNTDCNSNIIYVEAADVIIMHTPGCVPCVLEGSSRCWV 240
 DB 181 LLSCLTTPASAYEVNRVSGIYHVNTDCNSNIIYVEAADLIIMHTPGCVPCVLEGSSRCWV 240

QY 241 ALTPTLAARNASVPTTIRRHVDLLVGTAAAFCSAMYVDLCSIFIVSOLTFPSRRHET 300
 DB 241 ALTPTLAARNNSIPTTIRRHVDLLVGTAAAFCSAMYVDLCSIFIVSOLTFPSRRYET 300

QY 301 VQDCNCSLYPGHVSCHRMAMDMNWSPTTALVVSQLLRIPOAVVDMVAGAHGWLVLG 360
 DB 301 VQDCNCSLYPGHVSCHRMAMDMNWSPTTALVVSQLLRIPOAVVDMVAGAHGWLVLG 360

QY 361 YYSVGNWAKVLIIVALLFAGVDGTHHTGRVAGHTTSQFTSLFSSGASOKIQLVNTNGSW 420
 DB 361 YYSVGNWAKVLIIVLREFAGVDGTRVTVGAQVGRCTSSLTSIFTPGSPQNQLVNSNGSW 420

QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCRPIDWFAQGWGITVTKNSS 480
 DB 421 HINRTALSCNDSLTGFLAALFYTHKNASGCPERMASCRSIDTDFDQGWGITVHVPNT 480

QY 481 DQRPYCHWYAPRCGVVPASQVGVCTFSPVVGTTDRSGVPTYSWGENETDVLNL 540
 DB 481 DQPYCHWYAPRCGVVPASQVGVCTFSPVVGTTDRSGVPTYSWGENETDVLNL 540

QY 541 NTRPPQGNWFGCTWNSTGFTKTCGGPPCNIGGVGNRTLCPTDCFRKHPEATYTKCGSG 600
 DB 541 NTRPPQGNWFGCTWNSTGFTKTCGGPPCNIGGAGNNLTCTDCFRKHPEATYTKCGSG 600

QY 601 PWTLPCLVDYPYRLWHYPCTLNFSIFKVRMYVGVGVEHRLNAACNWRGERCNLEDRDRS 660
 DB 601 PWTLPCLVDYPYRLWHYPCTLNFTIFKIRMYVGVGVEHRLNAACNWRGERCNLEDRDRS 660

QY 661 ELSPLLSSTTEWQILPCAFITLTPALSTGLIHLHQNIVDVQYLYGVGSAFVFAIKWEYIL 720
 DB 661 ELSPLIRATTEWQILPCSFITLTPALSTGLIHLHQNIVDVQYLYGVGSAFVFAIKWEYIL 720

QY 721 LFLLLADARVCACLWMLLIAQAEALLENIVLNAASVAGAHGILSFLVFFCAAWYIKG 780
 DB 721 LFLLLADARVCACLWMLLIAQAEALLENIVLNAASVAGAHGILSFLVFFCAAWYIKG 780

781 RL 782
 781 RL 782

RESULT 14
 S18032
 genome polyprotein - hepatitis C virus (isolate JK4) (fragment)
 N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C;Species: hepatitis C virus
 A;Variety: isolate JK4
 C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C;Accession: S18032
 R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
 A;Reference number: S18029
 A;Accession: S18032
 A;Molecule type: genomic RNA
 A;Residues: 1-782 <HON>
 A;Cross-references: EMBL:J61594
 A;Experimental source: isolate JK4
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F;1-191/Product: core protein #status predicted <MAT1>
 F;192-383/Product: envelope protein 1 #status predicted <MAT2>
 F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 24.6%; Score 3942; DB 2; Length 782;
 Best Local Similarity 90.8%; Pred. No. 5.8e-232;
 Matches 710; Conservative 31; Mismatches 41; Indels 0; Gaps 0;

QY 1 MSTNPKPQRTKNTNRRPQDVKPGGQIVGGVYLLPRRGPRLGVRATRKASERSQPRG 60
 DB 1 MSTNPKPQRTKNTNRRPQDVKPGGQIVGGVYLLPRRGPRLGVRATRKATRSERSQPRG 60

QY 61 RROPPIKARRPEGRAWAQPQGYPMPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRSRNLG 120
 DB 61 RROPPIKARQPEGRACAPQGYPMPLYGNEGLGWAGWLLSPYGRPRMGPTDPRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATNGLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATNGLPGCSFSIFLLA 180

QY 181 LLSCLTTPASAYEVNRVSGIYHVNTDCNSNIIYVEAADVIIMHTPGCVPCVLEGSSRCWV 240
 DB 181 LLSCLTVPVSYEVNRVSGIYHVNTDCNSNIIYVEAADMIIMHTPGCVPCVREGSSRCWV 240

QY 241 ALTPTLAARNASVPTTIRRHVDLLVGTAAAFCSAMYVDLCSIFIVSOLTFPSRRHET 300
 DB 241 ALTPTLAARNNSIPTTIRRHVDLLVGTAAALCSAMYVDLCSIFIVSOLTFPSPRYET 300

QY 301 VQDCNCSLYPGHVSCHRMAMDMNWSPTTALVVSQLLRIPOAVVDMVAGAHGWLVLG 360
 DB 301 VQDCNCSLYPGHVSCHRMAMDMNWSPTTALVVSQLLRIPOAVVDMVAGAHGWLVLG 360

QY 361 YYSVGNWAKVLIIVALLFAGVDGTHHTGRVAGHTTSQFTSLFSSGASOKIQLVNTNGSW 420
 DB 361 YYSVGNWAKVLIIVLLFAGVDGTTVSGHSAQITRGVTSPSPGSAQIKIQLVNTNGSW 420

QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCRPIDWFAQGWGITVTKNSS 480
 DB 421 HINRTALNCNENITGFFAALFYANKFNSSGCSERMASCRPIDRFAQGWGITVHABSSS 480

QY 481 DQRPYCHWYAPRCGVVPASQVGVCTFSPVVGTTDRSGVPTYSWGENETDVLNL 540
 DB 481 DQRPYCHWYAPRCGVVPALQVGVCTFSPVVGTTDRFGAPTIVNNGANETDVLNL 540

QY 541 NTRPPQGNWFGCTWNSTGFTKTCGGPPCNIGGVGNRTLCPTDCFRKHPEATYTKCGSG 600
 DB 541 NTRPPQGNWFGCTWNSTGFTKTCGGPPCNIGVGNNKLTCTDCFRKHPEATYTKCGSG 600

QY 601 PWTLPCLVDYPYRLWHYPCTLNFSIFKVRMYVGVGVEHRLNAACNWRGERCNLEDRDRS 660
 DB 601 PWTLPCLVDYPYRLWHYPCTLVNFTIFKIRMYVGVGVEHRLNAACNWRGERCNLEDRDRS 660

QY 661 ELSPLLSSTTEWQILPCAFITLTPALSTGLIHLHQNIVDVQYLYGVGSAFVFAIKWEYIL 720
 DB 661 ELSPLLSSTTEWQILPCSFITLTPALSTGLIHLHQNIVDVQYLYGVGSAFVFAIKWEYIL 720

QY 721 LFLLLADARVCACLWMLLIAQAEALLENIVLNAASVAGAHGILSFLVFFCAAWYIKG 780

Db 721 LFLLLADARVACALWMLLIAQEAALLENVLVLAASVAGAHGVLPELVPFCAAWYIKG 780
QY 781 RL 782
Db 781 RL 782
RESULT 15
S19875
genome polyprotein - hepatitis C virus (isolate JK3) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK3
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S19875
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S19875
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61592; NID:G59482; PIDN:CAA43789.1; PID:G59483
A:Experimental source: isolate JK3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>
Query Match 24.5%; Score 3922; DB 2; Length 782;
Best Local Similarity 89.8%; Pred. No. 9.7e-231;
Matches 702; Conservative 37; Mismatches 43; Indels 0; Gaps 0;
QY 1 MSTNPKPQKTKRNTNRPDQVPPGGQIVGGVYLLPRGPRGLGVNATRKASRSQPRG 60
Db 1 MSTNPKPQKTKRNTNRPDQVPPGGQIVGGVYLLPRGPRGLGVNATRKASRSQPRG 60
QY 61 RRQPIKARPEGRARWQPGYFWPLYGNEGIGWAGWLLSPGSRPSPWGPDPFRSRNLG 120
Db 61 RRQPIKARPEGRARWQPGYFWPLYGNEGIGWAGWLLSPGSRPSPWGPDPFRSRNLG 120
QY 121 KVITLTCGPFADLMGYIPLVGAPLGGARALAHGVRLDGVNATGNLPCCSFSIFLLA 180
Db 121 KVITLTCGPFADLMGYIPLVGAPLGGARALAHGVRLDGVNATGNLPCCSFSIFLLA 180
QY 181 LLSCLITIPASAYERNVSGIYVHTNDGNSISIVYEADVMHTPGCVPCVGBGNSRCWV 240
Db 181 LLSCLITIPASAYERNVSGIYVHTNDGNSISIVYEADVMHTPGCVPCVGBGNSRCWV 240
QY 241 ALFTPLAARNASVPTTIRRHVDLLVGTAAFCASAMYVDLCGSIPLVSQSLTFPSRRHET 300
Db 241 ALFTPLAARNSSIPTTIRRHVDLLVGAALCSAMYVDLCGSIPLVSQSLTFPSRRHET 300
QY 301 VQDCNCSIYPGHVSGHRMAMDMANWSPTTALVVSQILRIPOAVVDMVAGAHGWGLAGLA 360
Db 301 VQDCNCSIYPGKLTGHRMAMDMANWSPTTALVVSQILRIPOAVVDMVAGAHGWGLAGLA 360
QY 361 YYSMVGNWAKVLIIVALLFAGVDGTHHTTGRVAGHTSGFTSLFSSGASOKIOLVNTNGSW 420
Db 361 YYSMVGNWAKVLIIVALLFAGVDGTHHTTGRVAGHTSGFTSLFSSGASOKIOLVNTNGSW 420
QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCREIDWFACQWGEITYTKPNSS 480
Db 421 HINRTAL-SCNDSLKTGFLAALFYTHKFNASGCPERMASCSRIDTDFDQGWGEITHVVENTT 480
QY 481 DQPYCHYARPCGVGPASQVGVGYCTPSPVVGTTDRSGV?TYSWGENETDVMNLN 540
Db 481 DQPYCHYARPCGVGPASQVGVGYCTPSPVVGTTDRSGV?TYSWGENETDVMNLN 540
QY 541 NTRPPQGNWFCCTWMSNGFTKTCGGPPCNTIGGVGNRTLICPTDCFRKHPEATYTKCGSG 600

Db 541 NTRPPQGNWFCCTWMSNGFTKTCGGPPCNTIGGVGNRTLICPTDCFRKHPEATYTKCGSG 600
QY 601 PWLTBRCCLVDYPYELWHYPCTLNFSIFKVRMYVGGVEHRLNAACNWRGRCNLEDRDRS 660
Db 601 PWLTBRCWVDYPYRLWHYPCTANFSVFKIRMYVGGVEHRLNAACNWRGRCNLEDRDRS 660
QY 661 ELSPLLLSTTTEWQILPCAFPTTLPALSTGLIHLHONIVDVQYLYXGVSFAVYFAIKWEYIL 720
Db 661 ELSPLLLSTTTEWQVLPCTFTTLPALSTGLIHLHONIVDVQYLYXGVSFAVYVVRWEYIL 720
QY 721 LIFILLADARVACALWMLLIAQEAALLENVLVLAASVAGAHGILSFLVFFCAAWYIKG 780
Db 721 LIFILLADARVACALWMLLIAQEAALLENVLVLAASVAGAHGILSFLVFFCAAWYIKG 780
QY 781 RL 782
Db 781 RL 782

Search completed: April 8, 2004, 10:49:46
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2004, 10:49:02 ; Search time 78 Seconds
(without alignments)
10146.708 Million cell updates/sec

Title: US-09-662-454-3

Perfect score: 16009

Sequence: 1 MSTNPKPQKTKNTNRKPQ.....FPLCLLLSVGVGHLLPNR 3010

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1073127 seqs, 262937947 residues

Total number of hits satisfying chosen parameters: 1073127

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14801.5	92.5	2985	14	US-10-259-275-40
2	14162.5	88.5	3011	9	US-09-742-659-4
3	14143.5	88.3	3011	10	US-09-891-894-3
4	14143.5	88.3	3011	14	US-10-184-150-3
5	14143.5	88.3	3011	15	US-10-328-997-3
6	14143.5	88.3	3012	9	US-09-238-076-2
7	14143.5	88.3	3012	10	US-09-995-937-2
8	14143.5	88.3	3012	10	US-09-917-563-2
9	14143.5	88.3	3011	12	US-10-326-724-406
10	14124.5	88.2	3011	9	US-09-922-572-9
11	14124.5	88.2	3011	9	US-09-747-419-20
12	14124.5	88.2	3011	12	US-10-189-359-14
13	14124.5	88.2	3011	14	US-10-259-275-20
14	14119.5	88.2	3011	9	US-09-916-359-2
15	14090.5	86.0	3011	9	US-09-238-076-20

16	14090.5	88.0	3011	10	US-09-995-937-20	Sequence 20, Appl
17	14090.5	88.0	3011	10	US-09-917-563-20	Sequence 20, Appl
18	13999.5	87.4	3011	14	US-10-232-643-6	Sequence 6, Appl
19	13935.5	87.0	3011	9	US-09-929-955-1	Sequence 1, Appl
20	13935.5	87.0	3011	13	US-10-104-966-1	Sequence 1, Appl
21	13442.5	84.0	2894	9	US-09-941-611-23	Sequence 23, Appl
22	13442.5	84.0	2894	14	US-10-044-995-23	Sequence 13, Appl
23	11614.5	72.5	2940	14	US-10-226-629A-13	Sequence 2, Appl
24	11145	69.6	2201	13	US-10-085-476-2	Sequence 3, Appl
25	11127	69.5	2201	13	US-10-029-907-3	Sequence 3, Appl
26	11127	69.5	2201	14	US-10-309-561-3	Sequence 3, Appl
27	10100	63.1	1985	14	US-10-259-275-42	Sequence 42, Appl
28	7143	44.6	1736	14	US-10-328-127-2	Sequence 2, Appl
29	7143	44.6	1736	14	US-10-328-206-2	Sequence 2, Appl
30	5887	36.8	2307	10	US-09-919-901-2	Sequence 2, Appl
31	5887	36.8	2307	14	US-10-191-966-2	Sequence 2, Appl
32	5884	36.8	2307	14	US-09-919-901-9	Sequence 9, Appl
33	5884	36.8	2307	14	US-10-191-966-9	Sequence 9, Appl
34	5875	36.7	2307	10	US-09-919-901-16	Sequence 16, Appl
35	5875	36.7	2307	14	US-10-191-966-16	Sequence 4, Appl
36	5862.5	36.6	1692	10	US-09-919-901-4	Sequence 4, Appl
37	5862.5	36.6	1692	14	US-10-191-966-4	Sequence 11, Appl
38	5859.5	36.6	1692	10	US-09-919-901-11	Sequence 11, Appl
39	5859.5	36.6	1692	14	US-10-191-966-11	Sequence 11, Appl
40	5850.5	36.5	1692	10	US-09-919-901-18	Sequence 18, Appl
41	5850.5	36.5	1692	14	US-10-191-966-18	Sequence 6, Appl
42	5326	33.3	2865	9	US-09-742-659-6	Sequence 9, Appl
43	4666.5	29.1	1040	9	US-09-929-955-9	Sequence 9, Appl
44	4666.5	29.1	1040	13	US-10-104-966-9	Sequence 9, Appl
45	4137	25.8	809	9	US-09-973-025-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1

US-10-259-275-40
; Sequence 40, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
; APPLICANT: Lencor, Stanley M.
; APPLICANT: Yi, Minkyung
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
; FILE REFERENCE: 265.0007 0120
; CURRENT APPLICATION NUMBER: US/10/259, 275
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US 60/171, 909
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 09/747, 419
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/325, 236
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/338, 123
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
; LENGTH: 2985
; TYPE: PRT
; ORGANISM: ARTIFICIAL
; FEATURE:
; OTHER INFORMATION: amino acid sequence encoded by nucleotides 2077-1121 of SEQ ID
; OTHER INFORMATION: NO:39
US-10-259-275-40

Query Match 92.5%; Score 14801.5; DB 14; Length 2985;

Best Local Similarity 91.5%; Pred. No. 0;

Matches 2764; Conservative 89; Mismatches 122; Indels 45; Gaps 5;

QY 1 MSTNPKPQKTKNTNRKPQDVFPGGQIVGGVLLPRGPRGLGVRAIKASERSQPRG 60

DB 1 MSTNPKPQKTKNTNRKPQDVFPGGQIVGGVLLPRGPRGLGVRAIKTKTSRQPRG 60

QY 61 RROPIKAREPEGRANAOQYPMPLXGNEGLGWAGWLLSPRGSRBWSGPTDPRRRNLG 120
 Db 61 RROPIKAREPEGRANAOQYPMPLXGNEGWAGWLLSPRGSRBWSGPTDPRRRNLG 120
 QY 121 KVIDTLTCGADLMGYIPLVGAALGGAARALAHGVRVLEDDGVNATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGLADLMGYIPLVGAALGGAARALAHGVRVLEDDGVNATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASAYEVNVSIGIHYHVNDCSNSIIVVBAADVIMHETPGVCPCVOEGNSRCWV 240
 Db 181 LLSCLTTPASAYEVNVSIGIHYHVNDCSNSIIVVBAADVIMHETPGVCPCVOEGNSRCWV 240
 QY 241 ALPPTLAARNASVPTTIRRHVDLLVGTAAFCSMYVGLCGSIFLYSOLFTSPRRHET 300
 Db 241 ALPPTLAARNASVPTTIRRHVDLLVGTAAFCSMYVGLCGSIFLYSOLFTSPRRHET 300
 QY 301 VQDCNCSIYPGHVSGHRMAMDMNMWSPPTALVVSQLLRIPQAVVDMVAGAHGVLGALA 360
 Db 301 VQDCNCSIYPGHVSGHRMAMDMNMWSPPTALVVSQLLRIPQAVVDMVAGAHGVLGALA 360
 QY 361 YYSNVGNWAKVILVALLFAGVDGETHTTGRVAGHTTSGFTSLPSSGASOKIQLVNTNGSW 420
 Db 361 YYSNVGNWAKVILVALLFAGVDGETHTTGRVAGHTTSGFTSLPSSGASOKIQLVNTNGSW 420
 QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCRPIDWFAQGWGPITYTKPNSS 480
 Db 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCRPIDWFAQGWGPITYTKPNSS 480
 QY 481 -DQPYCWHYAPRCVGVVPSQVCGPVYCTPSPVVVGTDRSGVPYVSGENBTDVMLL 539
 Db 481 PDQRPYCWHYPPQCGIIPVPSQVCGPVYCTPSPVVVGTDRSGVPYVSGENBTDVMLL 540
 QY 540 NNTPEPGNWFCTWNNSTGFTKCGGPPCNIGVGNRTLICPDCCPKRKEPATYTKGS 599
 Db 541 NNTPEPGNWFCTWNNSTGFTKCGGPPCNIGVGNRTLICPDCCPKRKEPATYTKGS 600
 QY 600 GPWLTPRCLVDYPRYLHYPCTLNFSIFKVENYVGGVEHRLNAACNTRGERCNLEDRDR 659
 Db 601 GPWLTPRCLVDYPRYLHYPCTLNFSIFKVENYVGGVEHRLNAACNTRGERCNLEDRDR 660
 QY 660 SELSPILLSTTEWQILPCAFPTLTPALSTGLIHLHQNIVDVOLYLVGVSFAFVSFAIKWEYI 719
 Db 661 SELSPILLSTTEWQILPCAFPTLTPALSTGLIHLHQNIVDVOLYLVGVSFAFVSFAIKWEYI 720
 QY 720 LLLPLLLADARVCAENMLLIARABALENLVNAAVAGAHGIIISFVPPCAANYIK 779
 Db 721 VLLFLLADARVCAENMLLIARABALENLVNAAVAGAHGIIISFVPPCAANYIK 780
 QY 780 GRAPGAAYAFYGVWPLLLLLALPPRAYALDREMAASCGGAVLVGLVFTLSYVYKVEL 839
 Db 781 GRAPGAAYAFYGVWPLLLLLALPPRAYALDREMAASCGGAVLVGLVFTLSYVYKVEL 840
 QY 840 TRLLWLOFYFTRABHMQWVWVPLNVRGGDAIILLTCAVHPELIIDITKLLAILGPL 899
 Db 841 ARLLWLOFYFTRABHMQWVWVPLNVRGGDAIILLTCAVHPELIIDITKLLAILGPL 900
 QY 900 MVLQAGITRVPYFVYAOGLIRACMLRVKAGHHVQVFMKIGALTGTYVYNHITPLRDW 959
 Db 901 MVLQAGITRVPYFVYAOGLIRACMLRVKAGHHVQVFMKIGALTGTYVYNHITPLRDW 960
 QY 960 AHAGRLDLAVAVEPVFSAMETKVIITWGAADTAACGDIIILGFPVSARRGKEIFLGPADSLE 1019
 Db 961 AHAGRLDLAVAVEPVFSAMETKVIITWGAADTAACGDIIILGFPVSARRGKEIFLGPADSLE 1020
 QY 1020 CQCHLLAPITAYSQOITRGVLCIITSLTGDKQVGEVQVSTATQSFATCINGCVW 1079
 Db 1021 CQCHLLAPITAYSQOITRGVLCIITSLTGDKQVGEVQVSTATQSFATCINGCVW 1080
 QY 1080 TVYHAGSKTLVAGKGPITQMYTNDLVGWQAPPGARSMTPCSCGSDLYLVTRADV 1139
 Db 1081 TVYHAGSKTLVAGKGPITQMYTNDLVGWQAPPGARSMTPCSCGSDLYLVTRADV 1140
 QY 1140 IPVRRRGDSRGLSPRPSVYLKSGSGGFLLCPSGHVGVFRAAVCTRGVAKAVDFIPE 1199

Db 1111 IPVRRGDNRSGLSPRPVSYLKGSSGGFLLCPSGHAVGVFAAVCTRGVAXAVDFVPE 1170
 QY 1200 SMETMRSVFTNDSTPPPAVQPTQVAHLHATGSGKSTKVPAAAYAAQYKVLVLPSPA 1259
 Db 1171 SMETMRSVFTNDSTPPPAVQPTQVAHLHATGSGKSTKVPAAAYAAQYKVLVLPSPA 1230
 QY 1260 ATIGFAYMSKAGIDPNIRTVGRTTTCGSIYSTYKFLADGGSGGAYDIICDECH 1319
 Db 1231 ATIGFAYMSKAGIDPNIRTVGRTTTCGSIYSTYKFLADGGSGGAYDIICDECH 1290
 QY 1320 STDSTTILGIVLQDAETAGARLVVLAATATPPGTVVPHPNIBETIGSNNGEIPFYGKA 1379
 Db 1291 STDSTTILGIVLQDAETAGARLVVLAATATPPGTVVPHPNIBETIGSNNGEIPFYGKA 1350
 QY 1380 IPIBATKGRHILFCHSKKKCBELAAKLTGLGNNAVAYVGLDVSVIPIGGVVVVATDA 1439
 Db 1351 IPTEATKGRHILFCHSKKKCBELAAKLTGLGNNAVAYVGLDVSVIPIGGVVVVATDA 1410
 QY 1440 LMTGFTGDSVLDNCNTVQVDFSLDPTFTTETTPQDAVRSQRGRGRTGRGSGTY 1499
 Db 1411 LMTGFTGDSVLDNCNTVQVDFSLDPTFTTETTPQDAVRSQRGRGRTGRGSGTY 1470
 QY 1500 RFVTPGERPSGMPDSSVLCCEYDAGCAWYELTPAETSVELRAYLNTFGLPVCQDHLFWE 1559
 Db 1471 RFVTPGERPSGMPDSSVLCCEYDAGCAWYELTPAETSVELRAYLNTFGLPVCQDHLFWE 1530
 QY 1560 SVPTGTHIDAHFSLSTKOAGNFPYLVAYQATVCARAQAPPSWDMWKKLIRLKPTLH 1619
 Db 1531 SVPTGTHIDAHFSLSTKOAGNFPYLVAYQATVCARAQAPPSWDMWKKLIRLKPTLH 1590
 QY 1620 GPTPLRYLGAVQNEVILTHPTIKYIMACMSADLEVVTVGLVGVLAALAAAYCITTS 1679
 Db 1591 GPTPLRYLGAVQNEVILTHPTIKYIMACMSADLEVVTVGLVGVLAALAAAYCITTS 1650
 QY 1680 VVTVGRILILSGKDAVVPDREVLVQEPDEMEBCASQLPYIEQGMOLABQKQKALGLQTA 1739
 Db 1651 VVTVGRILILSGKDAVVPDREVLVQEPDEMEBCASQLPYIEQGMOLABQKQKALGLQTA 1710
 QY 1740 TKQAEAAAPVVEKWEALETFWAKHWNFIISGQYLAGI-STLPGNPAIASLMAFTASITS 1799
 Db 1711 TKQAEAAAPVVEKWEALETFWAKHWNFIISGQYLAGI-STLPGNPAIASLMAFTASITS 1770
 QY 1800 PLTQNTLLPNIILGWWAQAQAPSAASAFAVAGIAGAAVGSIGLGLKVLDILAGYAGV 1859
 Db 1771 PLTQNTLLPNIILGWWAQAQAPSAASAFAVAGIAGAAVGSIGLGLKVLDILAGYAGV 1830
 QY 1860 AGALVAFKVMGSEVPSTEDLVNLLPALISFGALVGVVCAAILRRHVGPBGAVQWNNRL 1919
 Db 1831 AGALVAFKVMGSEVPSTEDLVNLLPALISFGALVGVVCAAILRRHVGPBGAVQWNNRL 1890
 QY 1920 TAPASRGNHVSPTHYVPESDAAARVTQIISLTLITQLKXELHOWINEDCSTPCSGSWLRD 1979
 Db 1891 TAPASRGNHVSPTHYVPESDAAARVTQIISLTLITQLKXELHOWINEDCSTPCSGSWLRD 1950
 QY 1980 VMDWICTVLSDFTKTLQSKLLPLPGVPLFSCQRYKGVWRGDMQITTCPCQAQIAGHV 2039
 Db 1951 VMDWICTVLSDFTKTLQSKLLPLPGVPLFSCQRYKGVWRGDMQITTCPCQAQIAGHV 2010
 QY 2040 KNGSMRLVGPRTCSNTWHGTFPINAITGCTPSPAPNYSRALWRVAABEYVEVTVGDF 2099
 Db 2011 KNGSMRLVGPRTCSNTWHGTFPINAITGCTPSPAPNYSRALWRVAABEYVEVTVGDF 2070
 QY 2100 HYVTGTMNVKPCQVPAPEFTEVDGVRLHRYAPACKPLLEDVTVFVGLNQYLVGSO 2159
 Db 2071 HYVTGTMNVKPCQVPAPEFTEVDGVRLHRYAPACKPLLEDVTVFVGLNQYLVGSO 2130
 QY 2160 L-----PCEPEPDVTVLSTMLTDPHSITAEAKRRLARGSPPSLASSASQISAPSLKAT 2214
 Db 2131 LHYRPAVCKP-----LLRDEVEVQVGLNQYLVGSOARGSPPSLASSASQISAPSLKAT 2185
 QY 2215 CTHH-----DSPADADILBANLLWRQMGNIITVSEKVKVILDSPEPLHAEDEREISV 2270

Db 2186 CTHSSYNLSDPDVDLJAAALLWQRMGGNIIRVESENKVVVLDSPEDLRABEDNEISI 2245
Qy 2271 AABILKSRKFPALPIWAPDYNPPLLESGWKDDPVVPVHGCPLPPTKAPDPPRRK 2330
Db 2246 AABILKSKKFPAAI2IWARPDYNPPLLESWKNDPVVPVHGCPLPPTKAPDPPRRK 2305
Qy 2331 RTVVLATESNVSSALAEIATKTFSGSSGSAVDSTATATPLASDDGDKGSDVESYSMP2 2390
Db 2306 RTVVLSDTSVSSVLAELATKTFGSELSSAADSGTATAPDQTDGDKGSDAESCSMP2 2365
Qy 2391 LEPEGDPDLSGWSVSEASESDVVCSSMSYTWGALITPCAABESKLPIPLNSLL 2450
Db 2366 LEPEGDPDLSGWSVSEASESVVCCSSMSYTWGALITPCAABESKLPIPLNSLL 2425
Qy 2451 RHNMVATTSRASLRQKVTDRIOLVDHHDVLEKMKAKASTVKAALLSIEACKL 2510
Db 2426 RHNMVATTSRASLRQKVTDRIOLVDHHDVLEKMKAKASTVKAALLSIEACKL 2485
Qy 2511 TPHSASKFGYGAQDVNLSRAVNHRSVWEDLIEDTETPDTTIMAKSEVFCVQPEK 2570
Db 2486 TPHSASKFGYGAQDVNLSRAVNHRSVWEDLIEDTETPDTTIMAKSEVFCVQPEK 2545
Qy 2571 GGRKPAELIVFPDLGVRVCEKXALYDVVSTLPQAVMGSSYGFQYSPQQRVEFLVWWSK 2630
Db 2546 GGRKPAELIVFPDLGVRVCEKXALYDVVSTLPQAVMGSSYGFQYSPQQRVEFLVWWSK 2605
Qy 2631 KCPMGFSYDTRCPDSTVETDIRVESIYQCCDLAPARQAISLTRELYIGGLTNSKG 2690
Db 2606 KCPMGFSYDTRCPDSTVETDIRVESIYQCCDLAPARQAISLTRELYIGGLTNSKG 2665
Qy 2691 QNCGYRCRAGSVLTTCGNTLCYKATACRAAKLQDCTMLVNGDDLVIICBSACTQE 2750
Db 2666 QNCGYRCRAGSVLTTCGNTLCYKATACRAAKLQDCTMLVNGDDLVIICBSACTQE 2725
Qy 2751 DAAALRAFTAMTRYSAPPDPOPEYDLELITSCSNVSVVAHDASGRVYLTDRPTTP 2810
Db 2726 DAAALRAFTAMTRYSAPPDPOPEYDLELITSCSNVSVVAHDASGRVYLTDRPTTP 2785
Qy 2811 LARAAMTARHTPNSWGLNIMVAPTLWARMILMTHTFETLLAOEQLKALDQIYGAC 2870
Db 2786 LARAAMTARHTPNSWGLNIMVAPTLWARMILMTHTFETLLAOEQLKALDQIYGAC 2845
Qy 2871 YSIEPLDLPQIIEHLGISAFTLSHSPGEINRVASCLRLKGVPPKRTWHRARSVRACL 2930
Db 2846 YSIEPLDLPQIIEHLGISAFTLSHSPGEINRVASCLRLKGVPPKRTWHRARSVRACL 2905
Qy 2931 LSQGRAATCRYLFFNAVRTKILITPPIAASOLDLSCNFWAGYSGGDIYHSLRARPRW 2990
Db 2906 LSQGRAATCRYLFFNAVRTKILITPPIAASOLDLSCNFWAGYSGGDIYHSLRARPRW 2965
Qy 2991 FPLCLLLSVGVGYLLPNR 3010
Db 2966 FPLCLLLSVGVGYLLPNR 2985

RESULT 2

US-09-742-659-4
; Sequence 4, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID0116
; CURRENT APPLICATION NUMBER: US/09/742,659
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22

; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-742-659-4
Query Match 88.5%; Score 14162.5; DS 9; Length 3011;
Best Local Similarity 85.7%; Pred. No. 0;
Matches 2579; Conservative 217; Mismatches 214; Indels 1; Gaps 1;
Qy 1 MSTNPKPQKTKENTNRPPQDVKPGGQIVGGVYLLPRGPRLGVRATRKZSESQPRG 60
Db 1 MSTNPKPQKTKENTNRPPQDVKPGGQIVGGVYLLPRGPRLGVRATRKZSESQPRG 60
Qy 61 RQPIPKARPPGRWAQPGYKPLVGNELGWAQWLLSPRGSRPSWGPTDPRRSRLUG 120
Db 61 RQPIPKARPPGRWAQPGYKPLVGNELGWAQWLLSPRGSRPSWGPTDPRRSRLUG 120
Qy 121 KVIDLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
Qy 181 LLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYEADVIMHTPGCVPCVQGNSSRCWV 240
Db 181 LLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYEADVIMHTPGCVPCVQGNSSRCWV 240
Qy 241 ALTPTLAARNASVPTTIRRHVDLNGTAAPCSAMVVGDLCSIFLVSQLFTPSRRHET 300
Db 241 AVTPTVATRDGKLPPTQLRRHIDLIVGSATLCSALVGDLCGSFVLVGLQFTPSRRHET 300
Qy 301 VDCNCSIYPGHVSHERMAWMMWNSPTALVVSQLLRIPOAVVDMVAGAHGVLVLAGLA 360
Db 301 TQCNCSIYPGHITGERMAMMMWNSPTAALVVAQLLRIPOAIMDMVAGAHGVLVLAGLA 360
Qy 361 YYSWGNWAKVLLVALLFAGVDGETHTRVAGHTTSFGTSLFSSGASQKIQLVNTNGSW 420
Db 361 YFSWGNWAKVLLVLLFAGVDAETHVTGGSAGHTTAGLVGLLTPGAKQIQLVNTNGSW 420
Qy 421 HINRTALNCDSIQTFEALFYAHKFNSSGCPERMASCREIDWFAGWGPIVYTKPNS 480
Db 421 HINSTALNCNESLNTGLWGLFTQHKFNSSGCPERLASCRLLTDFAGWGPIVYANGSGL 480
Qy 481 DORPYCWHYAPRGVVPASQVGVYCTFTSPVVGTTDRSGVPTYSGWENTDVMLLN 540
Db 481 DERPYCWHYAPRGVVPASQVGVYCTFTSPVVGTTDRSGVPTYSGWENTDVMVFLN 540
Qy 541 NTRPPQGNWFGCTWNNSTGTCTCGPPCMIGGVNRTLICPTDCPKHPEATVYKCGSG 600
Db 541 NTRPPGNWFGCTWNNSTGTCTKVGAPPVIGGVNNTLLCPTDCPKHPEATVYKCGSG 600
Qy 601 PMLTPRCIVDYPYRLWHYPTLNFSPKVMYGVGVHRLNACNMTGRERCNLEDRDRS 660
Db 601 PWITPRCWDYDYPHLYWHYPTCTINTYIFKVMYGVGVHRLNACNMTGRERCNLEDRDRS 660
Qy 661 ELSPLLLSTTQWVLPQCSFTTLPALSTGLIHLHONIVDVQVLYGVGSSIASMAIKWEYV 720
Db 661 ELSPLLLSTTQWVLPQCSFTTLPALSTGLIHLHONIVDVQVLYGVGSSIASMAIKWEYV 720
Qy 721 LFLLLADARVCACIWMMLLIAQAAALENLVNLNAAAGAHGILSFLVFFCAAWYIKG 780
Db 721 LFLLLADARVCSCILWMLLSQAAALENLVNLNAAAGAHGILSFLVFFCAAWYIKG 780
Qy 781 RLAPGAAVAFYGVWVPLLLILLALPPRAYALDREMAASCGGAVHVLVFLTSPYKVFELT 840
Db 781 RWPVGAVAFYGVWVPLLLILLALPPRAYALDREMAASCGGAVHVLVFLTSPYKVFELT 840
Qy 841 RLINWLOVFTIRAEAHQVWVPPVNVGGRDAIILLTCAVHPELIPDITKLLAILGLPLX 900
Db 841 WCMWWLQVFLTRVEAQLHVVVPEPLNVRGRDAVILLACVWHPTLVFDITKLLAILFGLW 900
Qy 901 VLQAGITRVPFVTRAQGLIRACMLRVKVGAGHYVQVFMKLGALTGTIVYNHILPLRDWA 960

[illegible]

RESULT 3
US-09-891-894-3


```

; Sequence 3, Application US/09891894
; Publication No. US2003003081A1
; GENERAL INFORMATION:
; APPLICANT: Olson, William
; APPLICANT: Madson, Paul
; TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITING HEPATITIS C VIRUS INF
; FILE REFERENCE: 2048/64896/JPW/SHS
; CURRENT APPLICATION NUMBER: US/09/891,894
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 3011
; TYPE: BRT
; ORGANISM: hepatitis c virus
;
US-09-891-894-3

Query Match      98.3%; Score 14143.5; DB 10; Length 3011;
Best Local Similarity 85.6%; Pred. No. 0;
Matches 2576; Conservative 218; Mismatches 216; Indels 1; Gaps 1;

Qy 1 MSTNPKXQRTKNTNRRPQDVKPGGGQIVGGVYLLPRRGPRLGVRAATKASRSQPRG 60
Db 1 MSTNPKXQRTKNTNRRPQDVKPGGGQIVGGVYLLPRRGPRLGVRAATKASRSQPRG 60

Qy 61 RQPIPKARREGRGAWAQPGWPPLYGNEG/GWAGWLLSPGSRPSWGPTDPRRRSRNLG 120
Db 61 RQPIPKARREGRGAWAQPGWPPLYGNEG/GWAGWLLSPGSRPSWGPTDPRRRSRNLG 120

Qy 121 KVIDTLTCGFADLMGYIPLVGAPLGGARALAHGVYVLEDGVNATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGARALAHGVYVLEDGVNATGNLPGCSFSIFLLA 180

Qy 181 LLSCLTTPASAYEVNVSIGIYHVTNDCNSISIVHEADVIWHTGCVPCVQBGNSRCWV 240
Db 181 LLSCLTTPASAYEVNVSIGIYHVTNDCNSISIVHEADVIWHTGCVPCVQBGNSRCWV 240

Qy 241 ALTPTLAARNASVPTTTRRVEDLLVGTAAFCSAMVGDLCGSLFLVSQLTFSPRRHET 300
Db 241 AVTFVATRDOKLPTTQLRHIDLLVGSATVCSALYVGDLCGSLFLVSQLTFSPRRHET 300

Qy 301 VQDCNCSIPYGHVSHRVAWMMNWSPTTALVVSQILRIQAVVDVWVAGAHGWLAGLA 360
Db 301 TQDCNCSIPYGHITGRVAVWMMNWSPTAALVVAQLLRIQAVINDMIAGAHGWLAGIA 360

Qy 361 YSMVGNWAKVLIYALLFAGVDGETHTRVAGHTTSFTSLFSGASOKTOLINTNGSW 420
Db 361 YFSMVGNWAKVLIYALLFAGVDAETHVTGGSAGRTTAGLVGLTPGAKONTOLINTNGSW 420

Qy 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCRPIDWFAQGWGPTTTPKNSS 480
Db 421 HINSTALNCNLSINTGWLAGLIFYQHKFNSSGCPERLASCRALTDFQAGWGPISYANGSGL 480

Qy 481 DQRPYCMHYAPRCGVVPASOVCGPVYCTFSPVYVVTGTTDSGVEPTYSWGNEDTVMLLN 540
Db 481 DERPYCMHYPRPCGVVPASOVCGPVYCTFSPVYVVTGTTDSGVEPTYSWGNEDTVMFVJN 540

Qy 541 NTRPPQGNKFGCTWMNSTGFTKTCGGPCNTGGVGNRTLIPTDCFRKHPEATYTKCGSG 600
Db 541 NTRPPLGNWFCGCTWMNSTGFTKTCGGPCNTGGVGNRTLIPTDCFRKHPEATYRSCGSG 600

Qy 601 PWLTPRCLVDYPRYLWHPYCTINPFIKRVYVGVGVEHRLAACNWTGRGERCNLZDRDRS 660
Db 601 PWITPRCVDYPRYLWHPYCTINPFIKRVYVGVGVEHRLAACNWTGRGERCNLZDRDRS 660

Qy 661 ELSPLLSTTEWQILPCFATPTLTPALSTGLIHJHQNIVDQXLYGVGSAFVSPAIKWEYIL 720
Db 661 ELSPLLSTTQWQLPCSFPTLTPALSTGLIHJHQNIVDQXLYGVGSSFASWAIKWEYV 720

Qy 721 LFLALLADARVCACIWMMLLIACRAALENVLVLAASVAGAHGTLFLVFFCAAWYIKG 780
Db 721 LFLALLADARVCSCLWMMLLISQBAALENVLVLAASLAGTHGLVSVLFFCFAYWIKG 780

```

```

781 RLAPCAAYAFYGVWPLLLLLLALPPRAYALDREMAASCGGAVLVGLVFLTLSPYKVFELT 840
781 RWPVGAAYAFYGVWPLLLLLLALPPRAYALDREMAASCGGAVLVGLVFLTLSPYKRYIS 840

841 RLITWQYFIFITRAEAMQVWVPLNVRGGRDALIILITCAVHPELIIDITKLLAILGLPM 900
841 WCWMLQYFLTRVEAQHLHWVPLNVRGGRDAVILLMVCVHPTLVFDITKLLAILGLPM 900

901 VLOAGITRVPYFVRQAQGLIRACMLVRKVAGGHVQVVMFKLGALTCTGYVYNHILTPLRDWA 960
901 ILQASLKLVPYFVRVQGLIRICALANKIAGCHVQVMAIILKGLTGTGYVYNHILTPLRDWA 960

961 HAGRLDAVAVEPVFSAMETKVITWGAADTAACGDIILGLPVSAARGKEIFIGPADSLEG 1020
961 HNGRLDAVAVEPVFSRMBETKLIITWGAADTAACGDIINGLPVSARRGQETILLGADGMVS 1020

1021 QGWRLLAPITAYSQOTRGVLGCIIITSLTGEDKNQVSEGVQVUSTATOSFLATCINGVCWT 1080
1021 KGWRLLAPITAYAQOTRGLGCIITSLTGEDKNQVSEGVQIVSTATQTFLATCINGVCWT 1080

1081 VYHGAGSKTLIAGPKGPTTOMYTNVDLDLVGWOAPPQGAARMTPCSCGSSDLVLYTRHADVI 1140
1081 VYHGAGRTTITASPKGPVIMYTNVDODLVGWPAQPSRSITPTCTCGSSDLVLYTRHADVI 1140

1141 PVRRRGDSRGLSLSPRPVSYLKGSSGGLLCPSGHVWVFPRAAVCTRGVAKAVDFIPVES 1200
1141 PVRRRGDSRGLSLSPRPISYLGKSSGGLLCPAGHAVGLFRAAAVCTRGVAKAVDFIPVEN 1200

1201 METTMSRPVTDNSTPPAVPQTVQVAHLHAPTSGSKSTKVPAAAYAGYKVLVLPNSVAA 1260
1201 LETTMSRPVTDNSTPPAVPQSFQVAHLHAPTSGSKSTKVPAAAYAGYKVLVLPNSVAA 1260

1261 TLGFGAYMSKAHGIDPNIRTVGRTITGGSTYTYGKFLADGCGSGAYDIIICDECHS 1320
1261 TLGFGAYMSKAHGVDPNIRTVGRTITGGSTYTYGKFLADGCGSGAYDIIICDECHS 1320

1321 TDSITLIGITVLDQAEATAGARLVWLNATAPPGSVVPHNIBIIGLSNNGEIPFYKAI 1380
1321 TDATSLIGITVLDQAEATAGARLVWLNATAPPGSVVPHNIBIIGLSNNGEIPFYKAI 1380

1381 PIEAIKGRHILRCHSKKCDLAAKLTGLGNNAVAYRGVLDVSVIPIPIGVVVVATDAL 1440
1381 PLEVIKGRHILRCHSKKCDLAAKLTGLGNNAVAYRGVLDVSVIPIPIGVVVVATDAL 1440

1441 MTGFTGDFSDVDCNTCTVQTVDFSLDPTTETTTVPQDAVSRSQRRGTGRGSIYR 1500
1441 MTGFTGDFSDVDCNTCTVQTVDFSLDPTTETTTVPQDAVSRSQRRGTGRGSIYR 1500

1501 FVTFGERPSGMFSSVLCYDAGCAWVELTPAETSVLRALAYLNTDGLPVQDHLFEWES 1560
1501 FVAFGERPSGMFSSVLCYDAGCAWVELTPAETSVLRALAYLNTDGLPVQDHLFEWEG 1560

1561 VFTGTLTHDAHFLSOTKQAGDNRPYLVAYQATVCARACAPPPSDQMKCLIRLKLPHG 1620
1561 VFTGTLTHDAHFLSOTKQSGENFPYLVAYQATVCARACAPPPSDQMKCLIRLKLPHG 1620

1621 PTPLLYRLGAYONEVILTHPTTKYIMACMSADLEWVTSTWLVGVVLAALAAAYCLTGSV 1680
1621 PTPLLYRLGAYONEVILTHPTTKYIMCMSADLEWVTSTWLVGVVLAALAAAYCLTGSV 1680

1681 VIVGRITLISGPVAVVPOREVLYQFDEMEBCASOLPIEQGMQLAEQPKQKALGLLOTAT 1740
1681 VIVGRITLISGPVAVVPOREVLYQFDEMEBCASOLPIEQGMQLAEQPKQKALGLLOTAS 1740

1741 KOEAAAAPVYESKRALETFWAKHMMNFISIGIYLAGLSTLPGNPALASLMAFTASTSP 1800
1741 KOEAVIPAVQTNCKLEVFVWAKHMMNFISIGIYLAGLSTLPGNPALASLMAFTASTSP 1800

1801 LTTQNTLLFNILGGVAAQAAPPSSAAGAFVAGIAGAAVSGISGLKVLVDILAGYGAGVA 1860
1801 LTTQNTLLFNILGGVAAQAAPPSSAAGAFVAGIAGAAVSGISGLKVLVDILAGYGAGVA 1860

1861 GALVAFKVMSEVSTEDLVNLLPAILSPGALVGVVCAAILRHHVGFEGGAVQWMMNELI 1920

```

1861 GALVAFKMGVSPSTEDLVNLLPALLSPGALVGVVCAILRRHVGPGEVQVWNRLL 1920
1921 AFASRGNHVSPTHYVPESDAARVTOILSSLTITQLKRLHOKWINEDECSPTSGGWLRDV 1980
1921 AFASRGNHVSPTHYVPESDAARVTOILSSLTITQLKRLHOKWINEDECSPTSGGWLRDV 1980
1981 KDWICVLTDEKTLQSLKLLPLPGVPEFLSCORGKGVWGRGDIMGTTCPCGAQIAGHVK 2040
1981 KDWICVLTDEKTLQSLKLLPLPGVPEFLSCORGKGVWGRGDIMGTTCPCGAQIAGHVK 2040
2041 NGSMRIVGPRTCENTWGTGTFPINAYTGTCTSPAPNYSRALMRVAAEYVYVTRVGDH 2100
2041 NGSMRIVGPRTCENTWGTGTFPINAYTGTCTSPAPNYSRALMRVAAEYVYVTRVGDH 2100
2101 YVTGMTDVKPCVPAPPEFTTEVDGVRHRYAPACKBLLEDVTFQVLNOLVGSOL 2160
2101 YVTGMTDVKPCVPAPPEFTTEVDGVRHRYAPACKBLLEDVTFQVLNOLVGSOL 2160
2161 PCPEPDVTLVSLMTPDPSHITAEAKRRLARGSPPLASSASQSLASPSLKATCTTHD 2220
2161 PCPEPDVTLVSLMTPDPSHITAEAKRRLARGSPPLASSASQSLASPSLKATCTTHD 2220
2221 SPDAELIANLWRQEMGNNITRVESENKVVILDSPEPLHAEGDEREISVAAEILKSKK 2280
2221 SPDAELIANLWRQEMGNNITRVESENKVVILDSPEPLHAEGDEREISVAAEILKSKK 2280
2281 FFSALDIWAPDPVNPBLLBSWKDPDVPVHVHCPPLPYKAPDIPRKRRTVLTESNV 2340
2281 FFSALDIWAPDPVNPBLLBSWKDPDVPVHVHCPPLPYKAPDIPRKRRTVLTESNV 2340
2341 SSALAMLATKTCGSSGSAVDSGTATPLDASDDGDKGSDVESYSMPLEGECDPDL 2400
2341 SSALAMLATKTCGSSGSAVDSGTATPLDASDDGDKGSDVESYSMPLEGECDPDL 2400
2401 SDGWSVTSEEA-SEDVCCMSYTGALITPCAAPESKLPINPLNSLLRHHNMVYAT 2459
2401 SDGWSVTSEEA-SEDVCCMSYTGALITPCAAPESKLPINPLNSLLRHHNMVYAT 2459
2460 TSASASLRQKVTDFRLOVLDHVRDLKEMKAKASTVAKLLSIREACKLTSPHSAKSK 2519
2461 TSASASLRQKVTDFRLOVLDHVRDLKEMKAKASTVAKLLSIREACKLTSPHSAKSK 2520
2520 FGKAKDVRNLSSRANVHRSVWEDLEDTETPIDTITMAKSEVFCVQPEKGRKPARLI 2579
2521 FGKAKDVRNLSSRANVHRSVWEDLEDTETPIDTITMAKSEVFCVQPEKGRKPARLI 2580
2580 VFDDLGVRVCEKVALYDVVSTLPOAVMGSSYGFQYSPKQVFEHLVNTWKSCKCPMGFSYD 2639
2581 VFDDLGVRVCEKVALYDVVSTLPOAVMGSSYGFQYSPKQVFEHLVNTWKSCKCPMGFSYD 2640
2640 TRCFDSVTESDIRVESIYOCODLAPEARQATSLRTERLYIGCLPLNSKGQCGVRRCR 2699
2641 TRCFDSVTESDIRVESIYOCODLAPEARQATSLRTERLYIGCLPLNSKGQCGVRRCR 2700
2700 ASGVLTSCGNTLTCYLKATAACRAAKLQDCTMLVNGDDLVWICESAGTQEDAAALRAFT 2759
2701 ASGVLTSCGNTLTCYLKATAACRAAKLQDCTMLVNGDDLVWICESAGTQEDAAALRAFT 2760
2760 EAMTRYAPGDDPQPEYDELELTSCSNVSVADHAGKRVYLTTRDPTPLAAAWETA 2819
2761 EAMTRYAPGDDPQPEYDELELTSCSNVSVADHAGKRVYLTTRDPTPLAAAWETA 2820
2820 RHTPINWLGNIIMYAPTLWARMILMTHFFSILLAOQLEKALDCOYIGACYSIEPDLDP 2879
2821 RHTPINWLGNIIMYAPTLWARMILMTHFFSILLAOQLEKALDCOYIGACYSIEPDLDP 2880
2880 QIIERLECLSAFTLHYSYSPGNEINRVAACLRKLGVPPLRTHRRARSVRKALLSOGGRAAT 2939
2881 PIIQRLHGLSAFSLHYSYSPGNEINRVAACLRKLGVPPLRTHRRARSVRKALLSOGGRAAT 2940
2940 CGRVLFWAVRTKLLTPIPAASOLDLSGNFVAGYSGGDIYHSLSRAPRWFPF-CILLLS 2999

2941 CGKYLFWAVRTKLLTPIPAASOLDLSGNFVAGYSGGDIYHSLSRAPRWFPF-CILLLS 3000
3000 VGVGYLLPNR 3010
3001 AGVGYLLPNR 3011
RESULT 4
US-10-184-150-3
; Sequence 3, Application US/10184150
; Publication No. US20030134297A1
; GENERAL INFORMATION:
; APPLICANT: Magdon, Paul
; APPLICANT: Magdon, Paul
; TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITING HEPATITIS C VIRUS INFI
; FILE REFERENCE: 2048/64896-A/JPM/MAF/DJK
; CURRENT APPLICATION NUMBER: US/10/184,150
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 09/891,894
; PRIOR FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 3011
; TYPE: PRN
; ORGANISM: Hepatitis C virus
US-10-184-150-3
Query Match 88.3%; Score 14143.5; DB 14; Length 3011;
Best Local Similarity 85.6%; Pred. No. 0; Mismatches 216; Indels 1; Gaps 1;
Matches 2576; Conservative 218;
QY 1 MGTNPKPKRKTNRNRRPQDVKPPGGQIVGGVLLPRRGLGVRAIRTKASERSQPRG 60
DB 1 MGTNPKPKRKTNRNRRPQDVKPPGGQIVGGVLLPRRGLGVRAIRTKASERSQPRG 60
QY 61 RQQTIPKARPRGAWAOPGYWPVLYNGLGAWGLLSPPGRSPSWPTDPRRSRLG 120
DB 61 RQQTIPKARPRGAWAOPGYWPVLYNGLGAWGLLSPPGRSPSWPTDPRRSRLG 120
QY 121 KVIDLITCGFADLMGVPLVGHPLGGAAPALAHGVVLEDDGVNATGNLPGCSFIFLLA 180
DB 121 KVIDLITCGFADLMGVPLVGHPLGGAAPALAHGVVLEDDGVNATGNLPGCSFIFLLA 180
QY 181 LLSCLTIPASAYEVRNVSGIYHVWINDCSNLSIVYEADVIMHTPGCVPCVQSGNSRCWV 240
DB 181 LLSCLTIPASAYEVRNVSGIYHVWINDCSNLSIVYEADVIMHTPGCVPCVQSGNSRCWV 240
QY 241 ALPTTAAARNASVPTTIRRHVDLLVGTAAFGSAMVYGLDCGSIHLVSQIFTPSPRHET 300
DB 241 ALPTTAAARNASVPTTIRRHVDLLVGTAAFGSAMVYGLDCGSIHLVSQIFTPSPRHET 300
QY 301 VODCNCSYIPGVSHRMAWDMAMWSPITALVVSOLLRIPOAVVDMAGAHGVLGLA 360
DB 301 VODCNCSYIPGVSHRMAWDMAMWSPITALVVSOLLRIPOAVVDMAGAHGVLGLA 360
QY 361 YXSMVGNWAKVILVALLFAGVDGETHITGRVAGHTTSGFTSLFSSGASOKIOLVNTGWS 420
DB 361 YXSMVGNWAKVILVALLFAGVDGETHITGRVAGHTTSGFTSLFSSGASOKIOLVNTGWS 420
QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCGPERMASCRPDWFAQGHPITYIKFNSS 480
DB 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCGPERMASCRPDWFAQGHPITYIKFNSS 480
QY 481 DQRPYCKWYABRPCGVVPASQVCGPVVCTPSPVVVGTTRDSCVPYSWGENETDVLIN 540
DB 481 DQRPYCKWYABRPCGVVPASQVCGPVVCTPSPVVVGTTRDSCVPYSWGENETDVLIN 540
QY 541 NTRPPOQNWPGCTWNNSTGFTKTCGGPPCPCNIGGVGNRTLICPTDCFRKPEPEYTKCGSG 600
DB 541 NTRPPOQNWPGCTWNNSTGFTKTCGGPPCPCNIGGVGNRTLICPTDCFRKPEPEYTKCGSG 600
QY 601 PWLTPTCLVDYPRLLWHYPCTLNFSIFPKVRMYVGVGVEHRLNAACTNTRGERCNLEDDRS 660

Db	601	WPIITPCWVDY	PVRLMHPCTINVTYIKVRMYGVGVEHLEAACNWTGRCDEDRDRS	660
Qy	661	ELSPLLISITTEWQILPCAFITLPAISLTGLHLHONIVDOYLYGVGSAPVSFAIKWEVIL	720	
Db	661	ELSPFLLSITTOWQVLPCSFITLPAISLTGLHLHONIVDOYLYGVGSSTIASMAIKWEVV	720	
Qy	721	LLFLLLLADARVACCLWNMLLIQAEEALENLVINAASVAGAHGILISLVFFCAAWYIKG	780	
Db	721	LLFLLLLADAAVCSCLWNMLLIISQAEALENLVILNAASLAGTHGLVSLVFFCFAWYILKG	780	
Qy	781	RLAPGAAYAFYGVWPLLILILILALPPRAYALDREMAASCGGAVILVGLVFTLTSPYKVFIL	840	
Db	781	RWVPGAAYAFYGVWPLLILILILALPQRAYALDTEVAASCGGVILVGLMALTLSPYKRYIS	840	
Qy	841	RLIILWLOYFITTRBAHNVQVWPLINVEGGRDAILILTCVAHPBELIDITKLLAILGPLM	900	
Db	841	WCMWMLQYFLTRVEAQJUEVWVPLINVRGGRDAILMLCVVPHFTLVFTITKLLAILGPLW	900	
Qy	901	VLQAGITRVPVFRVRAOGLIRACMLVRKVAGCHYVQVMFKLGALTITVYVNHILTPLRDA	960	
Db	901	ILQASLLKVPFVFRVQGLIRICALARKIAGHYVQMAIILKLGALTITVYVNHILTPLRDA	960	
Qy	961	HAGLRDLAAVPEVVPFGSMETKVITWGADTAAAGDIIILGLIPVSARKGEIFLGPADSLRG	1020	
Db	961	HNGLRDLAAVPEVVPFRMETKLTITWGADTAAAGDIINGLIPVSARRQETILGPADGMS	1020	
Qy	1021	QGWELLAPITAYSOOTRGVILCIIITSLTGDKNQVEGEVOVSTATQSLFATCINGCVMT	1080	
Db	1021	KGWELLAPITAYAOOTRGLLCIIITSLTGDKNQVEGEVOIVSTAQTFIATCINGCVMT	1080	
Qy	1081	VYHGAGSKTLAGPKGPITQMTYNVDLIVGWQAPPGARSMTPSCGSSDILYLVTREADVI	1140	
Db	1081	VYHGAGTETITASPKGPIVQMTYNVDQILVGPAPQGSRSITPTCTCGSSDILYLVTREADVI	1140	
Qy	1141	PVRRGDSRGSILSPRPVSYLKGSSGPELLCPCSGHVYGVFRAAVCTRGVAKAVDFTPVES	1200	
Db	1141	PVRRGDSRGSILSPRPISYLYKSGSGPELLCPCAGHAVGLFRAAVCTRGVAKAVDFTPVEN	1200	
Qy	1201	METTRSRPFTDNSTPPAIVPQTQVAVHIAHTPGSKSTKVPAAVAAQGYKVLVINSVAA	1260	
Db	1201	LETTNRSVPFTNSPSPAVPQSFQVHLEHAPTGGSKSTKVPAAVAAQGYKVLVINSVAA	1260	
Qy	1261	TLGFCAYMSKAHGDIPNIRTCVRIITGGSTITYYCKFLADGGCSGGAYDIIICDECHS	1320	
Db	1261	TLGFCAYMSKAHGVDPNIRTCVRIITGGSPITYYTYGKFLADGGCSGGAYDIIICDECHS	1320	
Qy	1321	TDSTTILGTVLDQAEITAGARLVVLATATPPGSVTVPHNIEBIGLNNGRIPFPYKAI	1380	
Db	1321	TDATSGILGTIVLDAQEITAGARLVVLATATPPGSVTVSHNIEIEVALSTTGEIPFPYKAI	1380	
Qy	1381	PLEAKTGRGHLIFCHSKKKCDLAALKTGLGLNAVAYYRGLDVSVIPIGDVVVVA'DAL	1440	
Db	1381	PLUEVTKGRHLIFCHSKKKCDLAALKVVALGINAVAYYRGLDVSVITSGDVVVVVA'DAL	1440	
Qy	1441	MTGFTGDFSDSDNCNTCVTQTVDFSLDPTFTTETITVPQDAVSRQSRGRTGRSGIYR	1500	
Db	1441	MTGFTGDFSDSDNCNTCVTQIVDFSLDPTFTTETITLQDAVSRQSRGRTGRGKGIYR	1500	
Qy	1501	FVTPGERSGMFDSSVLCECYDAGCAWVELTPAETSVALRAYLNTPGLPVCQDHLERWES	1560	
Db	1501	FVAPGERSGMFDSSVLCECYDAGCAWVELTPAETTVLRALRYMNTPLPVCQDHLERWEG	1560	
Qy	1561	VITGLTHIDAHFLSOTKQAGDNFPYLVAYQATVCARQAQAPPSDQMKCLIRLKP'TLHG	1620	
Db	1561	VFTGLTHIDAHFLSOTKSGENFPYLVAYQATVCARQAQAPPSDQMKCLIRLKP'TLHG	1620	
Qy	1621	PTPELLYELGAVQNEVILTHPIIKYIMACWSADLEVVITSTVVLGVGLAALAAAYCLITGCV	1680	
Db	1621	PTPELLYELGAVQNEVILTHPIIKYIMTCONSADLEVVITSTVVLGVGLAALAAAYCLITGCV	1680	
Qy	1681	VIVGRILISGKPAVVDREVLVQFDEMEBCASQYPIEQGQVLAEQFKQKALGLLO'ATAT	1740	

1681	VIVGRIVUSKPALIIDPREVLVOBFDREMBECSQLPIPIEQGMMLABQFKQKGLGLQTAS	1740
1741	QOAEAAAAPVYESKRALETFWAKHMWNFTSGYOYLAGJSTLPGNPATASLMAATASITSP	1800
1741	QOAEVITPAVQTNQWKLVEFWAKHMWNFTSGIYOYLAGJSTLPGNPATASLMAATAAVTSP	1800
1801	LTTQWTLIFNLGWAQAOLAPPSAASAFVGGAGIAGAAGVSGIGLKVLDILAGYGAGVA	1860
1801	LTTGQTLIFNLGWAQAOLAPGAATAFVGGAGIAGAAGVSGIGLKVLDILAGYGAGVA	1860
1861	GALVAFKVMGSEVPSTEDLVNLLPALILSPGALVGVWCAAILRRHVGPGBGAVQWNNRLI	1920
1861	GALVAFKVMGSEVPSTEDLVNLLPALILSPGALVGVWCAAILRRHVGPGBGAVQWNNRLI	1920
1921	AFASRGNHVSTHYVPESDMAARVTOILSSLIITOLLKRLHOWINEDCSTPCSGSMIRDV	1980
1921	AFASRGNHVSTHYVPESDMAARVTOILSSLIITOLLKRLHOWINEDCSTPCSGSMIRDI	1980
1981	WDKICTVLTDPKTMLOSKLPRILPGVPPFLSCOGYKGVWGRDGMQMOTPCPGACIAGHVK	2040
1981	WDWICEVLSDFKTLKAKLMPQLGPPFVSCORGYKGVWGRDGMQWTRCHGCAEITGHVK	2040
2041	NGSMRIVGPRCTSNWHTGTFPIINAYTTGCTPSPAKYSRALWRVAABEYEVTRVGDFFH	2100
2041	NGTMRIVGPRCTRNWNGTTFPIINAYTTGCTPLPAPNYKFALRVSAABEYEVTRVGDFFH	2100
2101	YVTGMTDNVXCPQVPAPEFFTEVDGVRILHRYAPACKPLJREBVTFOVGLNQVLVSQSL	2160
2101	YVSGMTDNLXCPQIPSPBFFTELDGVNLRHAFAPCKPLJREEVSFVGLHHEYFVQSOL	2160
2161	POEPPDPVTVLTSMLTDPSSHITAFTAKRLARGSPSSLASSASOLSAPSLKACTTHHD	2220
2161	POEPPDPVAVLTSMLTDPSSHITAAAGRLARGSPSSMASSASOLSAPSLKACTTANHD	2220
2221	SPDADILEANLLWQEMGNTTRVSESNKVILDSPEFLHABGDREELSAABELRKSXK	2280
2221	SPDAELILEANLLWQEMGNTTRVSESNKVILDSFDPLVAEDEKREYVPAELRKSXR	2280
2281	FPFSAPIWARPDYNPPLLESWKDDYVVPVHVHGCPLPPTKAPDIPPPRKRKTIVLTSNV	2340
2281	FAPAUPVWARPDYNPPLVETWKRPDYEPVHVHGCPLPPRSPVPPPRKKTIVLTESTL	2340
2341	SGALAEATKFGSGSGSANDVSGTATAPLPLASDDCKSDVESYSSMPLEGEFGPDPL	2400
2341	STALAEATKFGSGSSTSGITGNTTSSPAPSGCPCPSDVESYSSMPLEGEFGPDPL	2400
2401	SDGSGWSTVEEA-SDVVCSSMSYWTGALITPCAAEESKLPINPLNSLLRHHNMVYAT	2459
2401	SDGSGWSTVSSGADTEDVVCSSMSYWTGALVTFCAAEQKLPINALNSLLRHHNLVYST	2460
2460	TSRSASLRQKVTFRLOQLDDHYDRYLKBMKAKASTVAKULLSTIERACKUTPHESAKSK	2519
2461	TSRSACQRKQKVTFRLOQLDQVLDSHYODVLKVKAAAKVZANILLSVEEACSTPHESAKSK	2520
2520	FGYGAQVNRILSSRAVNHIRSWEDILEDTEPLDITTIMAKSEVFCVQPEKGGRRKPARLI	2579
2521	FGYGAQVNRCHARKVAHINSWDXLLEDSVTIDITTIMAKNEVFCVQPEKGGRRKPARLI	2580
2580	VFPDLGVRCERKALYDVVSTLPOAVNGSSYGFQYGPORVBFPLVNTWKSXKCPMGFSYD	2639
2581	VFPDLGVRCERKALYDVVSKLPLAVNGSSYGFQYSPGQRFVFLVQAKWSKKTPMGFSYD	2640
2640	TRCFDSTVTESDIRVEESIYQCCDIAPEARQAIRSLTERLYIGGPLTNSKGNOCYRRCR	2699
2641	TRCFDSTVTESDIRTEBAIYQCCDLDPQARVAIKSITERLYVGGPLTNSRGNGCYRRCR	2700
2700	ASGVLTTSCGNTLTCVYLKATACRAAKLODCTMLVNGDDLIVVICESAGTOEDAAALRAFT	2759
2701	ASGVLTTSCGNTLTCYIKARAACSAAGLQDCTMLVCGEDLIVVICESAGVQEDAAALRAFT	2760
2760	EAMTRYSAPGPBPQPEYDLELITSCSNNVSVAHDSKRVVYLTDPDTPPIKARAANWETA	2819
2761	EAMTRYSAPGPBPQPEYDLELITSCSNNVSVAHDGAKRVVYLTDRDTPPIKARAANWETA	2820

```

2820 RHTPINSWLGNIIMVAPTTLWARMILMTHFFSILLAQEOLEKALDQOIYGACYSIEPLDLP 2879
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2821 RUTPVNSWLGNIIMVAPTTLWARMILMTHFFSVLIARDQLEQALNIEIYGACYSIEPLDLP 2880
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2880 QIIEIRLHGLSAFTLHYSFGEINRVASCLRLKGVPLRTWRHRARSVRAKLLSQGGRAT 2939
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2881 PIIOQLHGLSATSLHYSFGEINRVAACTRLKGVPLRAWRHRARSVRARLLSRGGRAT 2940
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2940 CGRYLENWAVRTKLTPTPAASOLDLSCWEVAGYSGGDIYHSLSPARPRWFPLCLILL 2999
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2941 CGRYLENWAVRTKLTPTIAAGRLDLSGWTAGYSGGDIYHSVSHARPRWFPLCLLLA 3000
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3000 VGVGIYLLPNR 3010
      |||:|||||:
3001 AGVGIYLLPNR 3011
      |||:|||||:

RESULT 5
US-10-328-997-3
; Sequence 3, Application US/10328997
; Publication No. US20030232745A1
; GENERAL INFORMATION:
; APPLICANT: Olson, William
; APPLICANT: Maddon, Paul
; TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITING HEPATITIS C VIRUS INF
; FILE REFERENCE: 2048/64896-B/JPW/NAF/DJK
; CURRENT APPLICATION NUMBER: US/10/328,997
; CURRENT FILING DATE: 2002-12-24
; PRIOR APPLICATION NUMBER: 09/891,894
; PRIOR FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 10/184,150
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-328-997-3

Query Match 88.3%; Score 14143.5; DB 15; Length 3011;
Best Local Similarity 85.6%; Pred. No. 0;
Matches 2576; Conservative 218; Mismatches 216; Indels 1; Gaps 1;

QY 1 MSTNPKPQRTKNTNRPODVKFPGGQIVGGYILLPRGPRLGVRAATKASERSQPRG 60
DB 1 MSTNPKPQRTKNTNRPODVKFPGGQIVGGYILLPRGPRLGVRAATKASERSQPRG 60
QY 61 RRQIPKARPECEGRAWAQGYPMPLYNEGLGWAGWLLSPRGSRPWSMGPTDPRRRSRNLG 120
DB 61 RRQIPKARPECEGRAWAQGYPMPLYNECGGWAGWLLSPRGSRPWSMGPTDPRRRSRNLG 120
QY 121 KVITDLTCGFADLNGYIPLVGAPLGGAAALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVITDLTCGFADLNGYIPLVGAPLGGAAALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLITPASAYEVRNVSGIYHVTNDCNSSIYVEAADVIWHTPGCVPCVQEGNSSEFCW 240
DB 181 LLSCLITPASAYQVRNMSGIYHVTNDCNSSIYVEAADALHTPGCVPCVQEGNSASFCW 240
QY 241 ALPTTLAARNASVPTTTRRHVDLLVGTAAFCSAMVVGDLCSIFLVLSQLTFSPPRHET 300
DB 241 AVTETVATRGKLTPTQLRHIDLLVGSATLCSALYVGDLCGSFVLGVQLFTSPRHWT 300
QY 301 VQDCNCSITYPGHVSGHRMAMMMNWSPTTALVVSQLLRIPOAVVDMVAGAHGVLGAGLA 360
DB 301 TQDCNCSITYPCHITGHRMAMMMNWSPTTAAVVAQLLRIPQALMDMIAGAHGVLGAGLA 360
QY 361 YYSVWGNWAKVLIYALLIFAGVDGTHHTGVRAGHTTSGFTSLPSSGASQKIQLVNTGWS 420
DB 361 YFSVWGNWAKVLYVLLIFAGVDAETHVTGGSAGRTAGVGLLTPTGAKQNIQLNTGWS 420

```

1501 FNAPGRPSNFSSVLCYDAGCAYELTPBTTIVRLAAYNNTGLPCVQDHLRFWEG 1560
 1561 VFTGLTHIDAHFISQTKQAGDNFPYLVAQATVCARAQAPPPSDQWQKCLIRKPTLHG 1620
 1561 VFTGLTHIDAHFISQTKQAGDNFPYLVAQATVCARAQAPPPSDQWQKCLIRKPTLHG 1620
 1621 PTELLVRLGAVQNEVILTHPTIKYIMACMSADLEVVTSTWLVGGVLAALAAACLITGSV 1680
 1621 PTELLVRLGAVQNEVILTHPTIKYIMACMSADLEVVTSTWLVGGVLAALAAACLITGSV 1680
 1681 VIVGRITILSKPVPVDPREVLYQFEDDEMERCASOLPYIEQGMOLAEQFKQKALGLLQAT 1740
 1681 VIVGRIVLSKPAIPREVLYQFEDDEMERCASOLPYIEQGMOLAEQFKQKALGLLQAT 1740
 1741 KQAEAAAPVSVESKWRALETTWAKHMNFISGIVYAGLSTLPGNPATIASMTASTISP 1800
 1741 KQAEVITPAVQTNQKLEVFQWAKHMNFISGIVYAGLSTLPGNPATIASMTASTISP 1800
 1801 LTTQNTILFNILGWAQAOLAPPSAASAFYAGAGIAGAAVSGIGKVLVDILAGYGAV 1860
 1801 LTTQNTILFNILGWAQAOLAPPSAASAFYAGAGIAGAAVSGIGKVLVDILAGYGAV 1860
 1861 GALVAFKVMSEVSTEDLVNLLPAILSPGALVGVVCAALIRHRVQFEGGAVQWNRLLI 1920
 1861 GALVAFKIMSEVSTEDLVNLLPAILSPGALVGVVCAALIRHRVQFEGGAVQWNRLLI 1920
 1921 AFASRGNHVSPTHYVPSDAAARVQILSSITITOLLKRLHQWINECDSTPCSSWLRDV 1980
 1921 AFASRGNHVSPTHYVPSDAAARVQILSSITITOLLKRLHQWINECDSTPCSSWLRDV 1980
 1981 WDMTCVLTDFKWLQSLKRLPQVPELSCQGVKGVWRGDMOTTCPCGAOLAGHV 2040
 1981 WDMTCVLTDFKWLQSLKRLPQVPELSCQGVKGVWRGDMOTTCPCGAOLAGHV 2040
 2041 NGSMRIVGPRICSNTHGTFPINATTCPTSPAPNYSRALWRVAABEYVEVTRVGDH 2100
 2041 NGTWRIVGPICRNMWSTTFPINATTCPTPLPAPNYKFAWRVAABEYVEIRVGDH 2100
 2101 YVTGNTDNVPCQVPAPEFTEVDGVLRLHRYAPACKPLREDVTFQVGLNQYLVGSQL 2160
 2101 YVSGMTTDLNLCPCQIPSPFTELDGVLRLHRYAPACKPLREDVTFQVGLNQYLVGSQL 2160
 2161 PCEPDPVLTSMULTDPSHTAETAKSBLAGSPSSASSASQISAPSLKATCTTHD 2220
 2161 PCEPDPVLTSMULTDPSHTAETAKSBLAGSPSSASSASQISAPSLKATCTTHD 2220
 2221 SPDAELIEANLLWRQMGNTITRVESENKVVILDSFELHAECDREISVAAILRKSRK 2280
 2221 SPDAELIEANLLWRQMGNTITRVESENKVVILDSFELHAECDREISVAAILRKSRK 2280
 2281 FPSALPIWARPDYNPILASWKPDPYVPPVVGCPILPPTKAPPIPPRKRPTVLTESV 2340
 2281 FARALPVARPDYNPILPVTWKPPDYEPVWVGCPILPPTKAPPIPPRKRPTVLTESV 2340
 2341 SSALAEATKTPGSSGSAVDGTTALPDILASDGDGKSDVESYSSMPPLEGEPCDPL 2400
 2341 STALAEATKSGSSTSGITGNTTSSPPSPGCPDSDVESYSSMPPLEGEPCDPL 2400
 2401 SDGWSVTVEEA-SEBVVCCSSYTWGTGALITPCAAEESKLPINPLSNLLHHNNVTAT 2459
 2401 SDGWSVTSSGADTEDVCCSSYSYWTGALVTPCAAEEOKLPINALSNLLHHNNLVST 2460
 2460 TERSASLRQKVTDFELQVLDHYRVLKEMKASTVAKLISIEEACKLTPPHSAKSK 2519
 2461 TERSACOROKVTDFRLQVLDHYRVLKEMKASTVAKLISIEEACKLTPPHSAKSK 2520
 2520 FGAGKDVNLSRAVNHRSWVEDLEDETPTITIAKSEVFCVQPEKGRKPARLI 2579
 2521 FGAGKDVNRCHAKVAHNSVWKLLEDVTFIDITIAKSEVFCVQPEKGRKPARLI 2580
 2580 VFPDLGVRVCEKVALYDVVSTLFPQAVMGSSYGFQYSPKORVEFLVNTWKSCKPWFQFSYD 2639

2581 VFPDLGVRVCEKVALYDVVSTLFPQAVMGSSYGFQYSPKORVEFLVQAWKSCKTPMGFSYD 2640
 2640 TCECFSTVTSRIRVEESTYQCCDLAPARQAIRSLTERLYTGGPLTNSKGQNGYRRCR 2699
 2641 TCECFSTVTSRIRVEESTYQCCDLAPARQAIRSLTERLYTGGPLTNSKGQNGYRRCR 2700
 2700 ASGVLTSCTGNTLTCYKATAACRAAKIQDCTMLVNGDDLVI CESAGTQEDAAALRAFT 2759
 2701 ASGVLTSCTGNTLTCYKATAACRAAKIQDCTMLVNGDDLVI CESAGTQEDAAALRAFT 2760
 2760 EAMTRISAPGPPQPEYDLELITSCSSNVSAHADSGRRVYVLTDPPTPLARAWEA 2819
 2761 EAMTRISAPGPPQPEYDLELITSCSSNVSAHADSGRRVYVLTDPPTPLARAWEA 2820
 2820 RHTPINSWLNIIIMYAPTILWARMILMTHFFSIIIAQOELEKALDQOYGCYCSIEPLDLP 2879
 2821 RHTPINSWLNIIIMYAPTILWARMILMTHFFSIIIAQOELEKALDQOYGCYCSIEPLDLP 2880
 2880 QIIRLHGLSAFTLHSHSYSGEINRVASCLRKGVPPPLRTWRHRAARSVRKLLSQGGAAT 2939
 2881 PIQRLHGLSAFTLHSHSYSGEINRVASCLRKGVPPPLRTWRHRAARSVRKLLSQGGAAT 2940
 2940 CGYLFNWAVRTKLPTIPARASQDLDSGVFVAGYSGGDIYHLSRARPRWFPPLCILLIS 2999
 2941 CGYLFNWAVRTKLPTIPARASQDLDSGVFVAGYSGGDIYHLSRARPRWFPPLCILLIS 3000
 3000 VGVGIYLLPNR 3010
 3001 AGVGIYLLPNR 3011

RESULT 6
 US-09-238-076-2
 ; Sequence 2, Application US/09238076
 ; Patent No. US20020102540A1
 ; GENERAL INFORMATION:
 ; APPLICANT: RICE, CHARLES et al.
 ; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
 ; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
 ; STREET: 7733 FORSYTH BLVD., SUITE 1400
 ; CITY: ST. LOUIS
 ; STATE: MO
 ; COUNTRY: USA
 ; ZIP: 63105
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/238,076
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/034,756
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: HOLLAND, DONALD R.
 ; REGISTRATION NUMBER: 35,197
 ; REFERENCE/DOCKET NUMBER: 6029-4831
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 314-727-5188
 ; TELEFAX: 314-727-6092
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3012 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
US-09-238-076-2

Query Match 88.3%; Score 14143.5; DB 9; Length 3012;
Best Local Similarity 85.6%; Pred. No. 0;
Matches 2576; Conservative 2.8; Mismatches 216; Indels 1; Gaps 1;

QY	1	MSTNPQKQKTNTRRRQDKVFPQGGQIVGGVYLLPRGPRGLGVRAATKRSERSQPRG	60
DB	1	MSTNPQKQKTNTRRRQDKVFPQGGQIVGGVYLLPRGPRGLGVRAATKRSERSQPRG	60
QY	61	RQPIKARBPGRMAQPCQYMPYELGNEQLGAGWLLSPRGSRPBGWGTDPRRSRNLG	120
DB	61	RQPIKARBPGRMTWAQPCQYMPYELGNECGWAGWLLSPRGSRPBGWGTDPRRSRNLG	120
QY	121	KVDTLTTCGFADLMGYIPLVGAFLGGAARALAHGVRVLEGGVNYATGNLPGCCFSIFLLA	180
DB	121	KVDTLTTCGFADLMGYIPLVGAFLGGAARALAHGVRVLEGGVNYATGNLPGCCFSIFLLA	180
QY	181	LLSCLTIPASAEYEVNVSIGIHYHVNDCSNSSIVYEADVIMETPGCVPCVOECNSSRCMV	240
DB	181	LLSCLTIPASAYQVRNNSGIIHYVNDPCNSSIVYEADAHLHTPGCVPCVREGNASRCMV	240
QY	241	ALTEPTLAARNASVPTTIRRHVDLLVGTAAFCGAMVGDLCGSIFLVQSFLTFSPPRHET	300
DB	241	AVTPTVATRGKLTPTQLRRHIDLLVGSATLCSALYVGDLCGSVFLVQFLTFSPPRHWT	300
QY	301	VQDCNCSIIYPGHVSCHRMAMDMNMWSPPTALVVSQLLRIPQAVVDMVAGAHGWVLGALA	360
DB	301	TQDCNCSIIYPGHVSCHRMAMDMNMWSPPTAALVVAQLLRIPQAIMDMVAGAHGWVLGALA	360
QY	361	YYSVGVGWAKVLIYALLPAGVDGTHHTGRVAGHTTSGFTSLPSSGASQKIQLVNTNGSW	420
DB	361	YFSVGVGWAKVLIYALLPAGVDAETHVTGGSAGHTTAGLVGLTPGAKQNLQILNTNGSW	420
QY	421	HINTALNCDSLOTGFFAALFYAFKFNSSGCPERMASCRPIDWFAQGWGPITYTKPNSS	480
DB	421	HINSTALNCESLNTGLAGLFYQHKFNSGCPERLASCRRLTDFACQGWGPISYANGSL	480
QY	481	DQRPYCWHPYAPRPGCVGVASQVCGVCTTSPVVGTTDRSGVPTTYSWGENETDMLLN	540
DB	481	DERPYCWHPYPRPGCIYPAKSVCGPVYCTTSPVVGTTDRSGAPTSWGANDETDFVLN	540
QY	541	NTRPQGNWFCWTWNSTGFTKTCGGPECNLGGVGNRTLICPTCFRKHPEATYTKCGSG	600
DB	541	NTRPPLGNWFCWTWNSTGFTKVCAPPCVIGGVGNNTLLCPTCFRKHPEATYSRCSGS	600
QY	601	PWLTPRCLVDYPIYRLWHYFECTLNFSIFKVRVYVGVGVEHRLNAACNWTGERCNLEDRDS	660
DB	601	PWITPRCMVDYPIYRLWHYFECTINYITIFKVRVYVGVGVEHRLNAACNWTGERCDLEDRDS	660
QY	661	ELSPILLSTTZWQILPCAFITLIPALSTGLIHLHQNIVDVQVLYGVGSAFVSPAIKWEYIL	720
DB	661	ELSPILLSTTQVQLPCSFITLIPALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYV	720
QY	721	LFLLLADARVACALMMMLLTAQAEALLENVILNAAASVAGAHGILSPINFPCAAWYIKG	780
DB	721	LFLLLADARVACSLMMMLLISQAALLENVILNAAASLACTHGLVSLVFPFCAWYIKG	780
QY	781	RLAPGAAYAFYGVWVPLLLALLALPPRAYALDREMAASCGGAVLVGLVFLTSPYKVKVELT	840
DB	781	RWVPGAVAFYGVWVPLLLALLALPQRAYALDTEVAASCGGVVLVGLMALTLSPYKRVIS	840
QY	841	RLIWLQVPIITRABAMQVWVPLNVVRGGRDAIILLTCVHVELLFDITKILLAILGLIM	900
DB	841	WCMWMLQVPLTVEAQLHVVVPLNVVRGGRDAVILLCMVHVELLFDITKILLAILFGLW	900
QY	901	VLOAGITRVPYVRAQGLLRACMLKVRKAGGHYVQVMFKGALGTYYVNHILTPLRDWA	960
DB	901	ILQASLLKPYFVRVQGLLURIICALARKIAGGHYVQVMAILKUGALGTYYVNHILTPLRDWA	960
QY	961	HAGLRDLAVAVEFVVFSAKETVITWGADTAACGDIILGLPVSARRGKEIFLGPADSLBG	1020

DB	961	HNGLRDLAVAVEFVVFSEMETKLITWGADTAACGDIINGLPVSARRGOEIIILGPADGHVS	1020
QY	1021	QGWRLLAIPITAYSQOTRGVLGCIITSLTGRDKNOVEGEVQVWVSTATQISFLATCINGVCWT	1080
DB	1021	KWRLLAIPITAYAQOTRGLGCIITSLTGRDKNOVEGEVQIVSTATQIFLATCINGVCWT	1080
QY	1081	VYHGAGSKTLAPGKPIQMYTNVDLDJVGWQAPPGARSWTPCSCGSDILYLVRHADVI	1140
DB	1081	VYHGAGTATIASPKGEVIQMYTNVDQDLVGWPAPOGSRSLTPTCGSDILYLVRHADVI	1140
QY	1141	PVRRGDSRGSLLSRPVSYLKGSGGGLCPESHGHWGVFAAAVCTRGVAKAVPIPVES	1200
DB	1141	PVRRGDSRGSLSRPISYLYKGSGGGLCPAGHAGVLFRAAVCTRGVAKAVPIPVEN	1200
QY	1201	METTWRSPVFTDNSTPPPAVPQTFQVAHLHAPTGSCKTKVPAAYAAQYKVLVLPNSVAA	1260
DB	1201	LETTWRSPVFTDNSSPPPAVPQSFQVAHLHAPTGSCKTKVPAAYAAQYKVLVLPNSVAA	1260
QY	1261	TLGFGAYMSKAHGDIPNIRNTGVRTITTTGGSITYSYGKFLADGGCGSGGAYDIIICDCHS	1320
DB	1261	TLGFGAYMSKAEGVDPNIRNTGVRTITTTGSPITYSYGKFLADGGCGSGGAYDIIICDECHS	1320
QY	1321	TDSITILGIGTVDQAEATAGALVLAATATPPGSVTVPHENIEEICLSNNGEIPYVGKAI	1380
DB	1321	TDATSILGIGTVDQAEATAGALVLAATATPPGSVTVSHENIEEVALSTITGELPFYVGKAI	1380
QY	1381	PTIAIKGGRHLLFCHSKKCKDELAALKJGLGLNNAVAYRGLDVSVIPIPIGDVVVWATDAL	1440
DB	1381	PLEVIKGRHLLFCHSKKCKDELAALKVALGINAVAYRGLDVSVIPTSVDVVWSTDAL	1440
QY	1441	MTGFTGDFDSVIDCNTCVTQVDFSLDPIPTTITTTVPQDANVSORSORCGRGSIYR	1500
DB	1441	MTGFTGDFDSVIDCNTCVTQVDFSLDPTFTTITTLPODANVSORTORRGRGKEIYR	1500
QY	1501	FVTPGSRBSPMPDSSVLCEYDAGCAWYELTDAETSRLRAYLNTPLGLPVCDHLEFWS	1560
DB	1501	FVAPGSRBSPMPDSSVLCECYDAGCAWYELTDAETTVRLRAYNLTGLPVCDHLEFWS	1560
QY	1561	VFTGLTHIDAHFLSQTKAGDNFPYLVAYQATVCARACAPPPSWDQWCKLIRLKLTHG	1620
DB	1561	VFTGLTHIDAHFLSQTKSGENFPYLVAYQATVCARACAPPPSWDQWCKLIRLKLTHG	1620
QY	1621	PTLLVRLGAVQNEVILTHPIKTYIMACMSADLEVVTSTWLVGGVLAALAAAYCLLTGCV	1680
DB	1621	PTPLLYRLGAVQNEVILTHPIKTYIMTMSADLEVVTSTWLVGGVLAALAAAYCLLTGCV	1680
QY	1681	VIVGRITLSCKPAVDPREVLYOEFEDEMERECASQLPYIEQGMOLABQFKOKALGLQTAT	1740
DB	1681	VIVGRIVLSKPAIIPREVLYOEFEDEMERECOSHLPYIEQGMMLABQFKOKALGLQTAS	1740
QY	1741	KQABAAAAPVVSKEWRALETFWAKEMNFIISGIIYLAGLSTLPGNPAIASIMAFASITSP	1800
DB	1741	RQABVITPAQVNWQKLEVPWAKEMNFIISGIIYLAGLSTLPGNPAIASIMAFATAVTS	1800
QY	1801	LTQNTQLLFNILGWNVAQIAPSAASAFVAGTAGAVSGIGLVKVLVDILAGYAGVA	1860
DB	1801	LTTCQTLLFNILGWNVAQIAPGAATFVAGTAGIAGIISVGLKVLVDILAGYAGVA	1860
QY	1861	GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVVGVCVCAAILRRHVGPBGAVQWNRLLI	1920
DB	1861	GALVAFKIMGSEVPSTEDLVNLLPAILSPGALVVGVCVCAAILRRHVGPBGAVQWNRLLI	1920
QY	1921	AFASRGHNVSPTHVVPESDAAARVTQILSLITITQLLKRHLQHNINEDCSTPCSGSWLRDI	1980
DB	1921	AFASRGHNVSPTHVVPESDAAARVTAILSLITVQLLRLLQHNINESCTTPCSGSWLRDI	1980
QY	1981	WDWICTVLTDPKTMQSKLLPLPGVFLSCORYKGVWRCGDMQTTCPGAGIAGHVK	2040
DB	1981	WDNICEVLSDPKTKAKMLPQLECIPIVSCQRYGVWRCGDMGVYHTRCHGAGIAGHVK	2040
QY	2041	NGSMELVGPCCSNWGTGTFPINAVTTGCTCPSPAPNYSRALMEVAAAEYEVTRVGDH	2100

2041 NGTMEIVGPRTCRNWSGTPIINAYTTGPTCTPLPAPNYKFALEMSAEYVEIRRVGDPH 2100
 2101 YVTGMITDNDVCKPCQVPAPEFFTEVDGVLRLHRYAPACKPLIREDDVTQVGINQVIVGSQL 2160
 2101 YVSGMITDNLKPCQIQIPSEFFTELDGVLRLHRYAPACKPLIREDDVTQVGINQVIVGSQL 2160
 2161 PCPEPDDVTILTSMLTDPESHITATKRRRLARGSPPLASSASASQLSAPSLKATCTTHHD 2220
 2161 PCPEPDDVAVILTSMLTDPESHITABAGRLARGSPPLASSASASQLSAPSLKATCTANHD 2220
 2221 SPDADLIEANLLWQMGNTTRVESNKVVILSPFPLHAGDEREISVAEILRLSRK 2280
 2221 SPDADLIEANLLWQMGNTTRVESNKVVILSPFPLHAGDEREISVAEILRLSRK 2280
 2281 FPSALP-WARPDYNNPPLLESKDDYPPVVGCGPLPPTKAPPIPPRRKRTVIVTESNV 2340
 2281 FARALPWARPDYNNPPLLESKDDYPPVVGCGPLPPTKAPPIPPRRKRTVIVTESNV 2340
 2341 SSALAEATKTFGSGSSAVDSGTATLPDLASDGDGKGSVESYSMPPLGEPGDPDL 2400
 2341 STALAEATKTFGSGSSAVDSGTATLPDLASDGDGKGSVESYSMPPLGEPGDPDL 2400
 2401 SDGKSTVSEBA-SBDVCCSNSTWTGALITPCAAEESKIPINPLNSLLRHNNMYAT 2459
 2401 SDGKSTVSSGADTDVCCSNSTWTGALITPCAAEESKIPINPLNSLLRHNNMYAT 2460
 2460 TSRSASLRQKVTDFRLOVLDHVDVLEKEMKAKASTVKALLS-EEACKLTPPHSAKSK 2519
 2461 TSRSACQKQKVTDFRLOVLDHVDVLEKEMKAKASTVKALLS-EEACKLTPPHSAKSK 2520
 2520 FGYGAKDVNLSRAVNHRSVWEDLDTPTIDTTIMAKSEVFCVQPERGGRKPARLI 2579
 2521 FGYGAKOVKCHARKAVAHINSVWDLDSVTPIDTTIMAKNEVFCVQPERGGRKPARLI 2580
 2580 VPPDLGVRVCVMALYDVVWILPQAVMGSSYGFQVSPQRVEFLVNTWKSXKCPMGFSYD 2639
 2581 VPPDLGVRVCVMALYDVVWILPQAVMGSSYGFQVSPQRVEFLVNTWKSXKCPMGFSYD 2640
 2640 TRCPDSTVTSRDIRVERSIYOCCLAPARAOAIRSLTERLYIGGLTNSKGNOCYRRCR 2699
 2641 TRCPDSTVTSRDIRVERSIYOCCLAPARAOAIRSLTERLYIGGLTNSKGNOCYRRCR 2700
 2700 ASGVLTSCGNTLCYLKATACRAAKLQDCTMLVNGDDLWVICSAGTQEDAAALRAFT 2759
 2701 ASGVLTSCGNTLCYLKATACRAAKLQDCTMLVNGDDLWVICSAGTQEDAAALRAFT 2760
 2760 EAMTRYAPPDPPQPEYDLELITSCSNVSVAHDASGRVYVYLTDRPTTLARAAMETA 2819
 2761 EAMTRYAPPDPPQPEYDLELITSCSNVSVAHDASGRVYVYLTDRPTTLARAAMETA 2820
 2820 RHTPNSWLGNTIMVAPTLWARMILMTHTFFSILLAOEOLKALDCQIYGACYSIEPLDIP 2879
 2821 RHTPNSWLGNTIMVAPTLWARMILMTHTFFSILLAOEOLKALDCQIYGACYSIEPLDIP 2880
 2880 QIIEHLGLSAPTLHSYSGEINRVASCLKUGVPLPRTWHRARSVRKLLSQGGRAT 2939
 2881 QIIEHLGLSAPTLHSYSGEINRVASCLKUGVPLPRTWHRARSVRKLLSQGGRAT 2940
 2940 CGKYLFNNAVTKLKTPIPAASQDLSCGFVAGYSGGDIYHLSRARPWFPLCLILL 2999
 2941 CGKYLFNNAVTKLKTPIPAASQDLSCGFVAGYSGGDIYHLSRARPWFPLCLILL 3000
 3000 VGVGIYLLPNR 3010
 3001 AGVGIYLLPNR 3011

RESULT 7

US-09-995-937-2
 ; Sequence 2, Application US/0995937
 ; Publication No. US20030028010A1
 ; GENERAL INFORMATION:
 ; APPLICANT: RICE, CHARLES et al.

TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
 VIRUS (HCV) AND USES THEREOF
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: HOWELL & HAFERKAMP, L.C.
 STREET: 7733 FORSYTH BLVD., SUITE 1400
 CITY: ST. LOUIS
 STATE: MO
 COUNTRY: USA
 ZIP: 63105
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/995,937
 FILING DATE: 28-Nov-2003
 CLASSIFICATION: <UNKNOWN>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/034,756
 FILING DATE: 04-May-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 REFERENCE/DOCKET NUMBER: 6029-4831
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 314-727-5188
 TELEFAX: 314-727-6092
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3012 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 FRAGMENT TYPE: N-terminal
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-995-937-2

Query Match 88.3%; Score 14143.5; DR 10; Length 3012;
 Best Local Similarity 85.6%; Pred. No. 0;
 Matches 2576; Conservative 218; Mismatches 216; Indels 1; Gaps 1;
 Qy 1 MSTNPKPORKTKRNTRRRPODVKPPGGQIVGGVYLLPRRGRPLGVRATRKASERSQPRG 60
 Db 1 MSTNPKPORKTKRNTRRRPODVKPPGGQIVGGVYLLPRRGRPLGVRATRKASERSQPRG 60
 Qy 61 RRQPIPKARREPEGRAWAQPGYPWLYGNELGHWAGWLLSPRGRSPSGPTDPRRSRLG 120
 Db 61 RRQPIPKARREPEGRAWAQPGYPWLYGNELGHWAGWLLSPRGRSPSGPTDPRRSRLG 120
 Qy 121 KVIDTLTCGFDIADMGYPLVGCAPLGGARALAHGVLEEDGVNATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFDIADMGYPLVGCAPLGGARALAHGVLEEDGVNATGNLPGCSFSIFLLA 180
 Qy 181 LLSCLTIPASAYEVNRSVGIYHVINDCSNNSIVVEAADVIMETPCVPCVQEGNSSRCWV 240
 Db 181 LLSCLTIPASAYEVNRSVGIYHVINDCSNNSIVVEAADVIMETPCVPCVQEGNSSRCWV 240
 Qy 241 ALPTPLAARNASVPTTTRRHVDLLVGTAAAPCSAMVYVGDLCGSLFVLSQLTFFSPRHET 300
 Db 241 AVTPTVATRDGKLPPTQLRRHIDLLVGSATLCSALYVGDLCGSLFVLSQLTFFSPRHET 300
 Qy 301 VQDNCNSIYEGHVSCHRMAMWMMNNSPTTALVVSQILRIPOAVVDMVAGAHGWVGLA 360
 Db 301 TQDNCNSIYEGHVSCHRMAMWMMNNSPTTALVVSQILRIPOAVVDMVAGAHGWVGLA 360
 Qy 361 YFSVMGNWAKVLIIVALLPAGVDGETHTTGRVAGHTTSGFTSLFSSGASQKIQLVNTNGSW 420
 Db 361 YFSVMGNWAKVLIIVALLPAGVDGETHTTGRVAGHTTSGFTSLFSSGASQKIQLVNTNGSW 420

QY 421 HINRTALNCNDSIQTGEAALFYAHKFNSSGCGPERMASCRIPIIDWFAQGWPIITYKPNSS 480
 Db 421 HINSTALNCNESLNTGWLAGLFFQHKFNSSGCGPERLASCRRLTDFQAQGWPIISYANGSGL 480
 QY 481 DORPYCWHYAPRCGVVVPASQVCGPYCYFTPSVYVGTTRDSGVPTPYSWGNETDVMNLN 540
 Db 481 DESPYCWHYPPRCGIVFPAKSVCGPVYCTPSVYVGTTRDSGAPYSGANDTDVFIN 540
 QY 541 NTRPPQGNFGCTWMSGTFTKCGPPCNIGVGNRTLICPTDORRKHPEATYTKCGSG 600
 Db 541 NTRPPQGNFGCTWMSGTFTKCGPPCNIGVGNRTLICPTDORRKHPEATYTKCGSG 600
 QY 601 PWLTPTCLNDYVYRLWHYPCTLNFSIFKVMYVGVGVEHRLNAACTRGERCMLEDRDS 660
 Db 601 PWLTPTCLNDYVYRLWHYPCTLNFSIFKVMYVGVGVEHRLNAACTRGERCMLEDRDS 660
 QY 661 ELSPLILLSTTEWQILPCAFPTLLPALSTGLIHLHONIVDVQYLYGVGSAFVSPFAIKWEYIL 720
 Db 661 ELSPLILLSTTEWQILPCAFPTLLPALSTGLIHLHONIVDVQYLYGVGSAFVSPFAIKWEYIL 720
 QY 721 LLELLILADAEVCACLNMWMLIIAABALENIVLNAASVAGAGIISFELVFPCCAAMVYK 780
 Db 721 LLELLILADAEVCACLNMWMLIIAABALENIVLNAASVAGAGIISFELVFPCCAAMVYK 780
 QY 781 RLAPGAAYAFYGVMPILLILLALLPPRAYALDRMAASCGGAVLGVHFLTLSPYKVFLT 840
 Db 781 RLAPGAAYAFYGVMPILLILLALLPPRAYALDRMAASCGGAVLGVHFLTLSPYKVFLT 840
 QY 841 RLILWLOYFTTRABAEQVWVPELNVBGRDADIILLTCAVHPELIIIDITKLLAILLIGPLM 900
 Db 841 WCMWLOVFLTRVBAQJHVWVPELNVBGRDADIVLLMCVHPTLVDFDITKLLAILLIGPLM 900
 QY 901 VLOAGITRVPYVBAQGLACMLVVRVAGVHYVQVPMKLGALTGYVYNHLLTPLRDWA 960
 Db 901 ILQASLKLVPFVVRQGLRICALARKIAGHYVQMAIILKLGALTGYVYNHLLTPLRDWA 960
 QY 961 HAGLRDLAVAVPEVYVFSAMETKVIWGAADTAACGDIILGLPVSARRKETFLGPADSLEG 1020
 Db 961 HNGLRDLAVAVPEVYVFSAMETKVIWGAADTAACGDIILGLPVSARRQEIILGPADGWS 1020
 QY 1021 QGWKLLAPITAYSQOTRGVLGCIITSLTGRDKNOVEGEVQVWSTATOSFLATCINGCWT 1080
 Db 1021 KGWKLLAPITAYAOQTRGLLGCIITSLTGRDKNOVEGEVQVWSTATQTLATCINGCWT 1080
 QY 1081 VYHAGSKTLAGPKGPIQMTNVNDLIVGQAAPPGARSTPCCSGSDLYLVTRHADVI 1140
 Db 1081 VYHAGGTRTASPKGPVQMTNVNDQLVGPAPQGSRLTPTCTCGSDLYLVTRHADVI 1140
 QY 1141 PVRREGDSRGLLSPRPVSYLKGSGGPLLCPSGHVGVFPAAVCTRGVAKAVDFIPVES 1200
 Db 1141 PVRREGDSRGLLSPRPISYLKGSGGPLLCPAGHAVGLPEAAVCTRGVAKAVDFIPVEN 1200
 QY 1201 MEITWRSVPFTDNTSPRAVPOTFQVAHLHAPTGSCKTKVPAAYAAQYKVLVLPNSVAA 1260
 Db 1201 LETTWRSPVFTDNTSPRAVPOSFQVAHLHAPTGSCKTKVPAAYAAQYKVLVLPNSVAA 1260
 QY 1261 TLGFGAYMSKAGHDIPNRTGVRTITGGSITYSTYKFLADGGCGGAYDIILICDECHS 1320
 Db 1261 TLGFGAYMSKAGHDVPNRTGVRTITGGSPIITYSTYKFLADGGCGGAYDIILICDECHS 1320
 QY 1321 TDSVTILGIVLVDOAETAGARLVVLATATPPGSVTVVPHNIEBGLSNNEIIPYKAI 1380
 Db 1321 TDAVSILGIVLVDOAETAGARLVVLATATPPGSVTVVSHNIEEVALSTTGEIIPYKAI 1380
 QY 1381 PIEATKGRSHLIFCHSKKKCBELAKITCLGINAVAYVREGLDVSIPIPGDVVVVATDAL 1440
 Db 1381 PLEVTKGRSHLIFCHSKKKCBELAKI VALGINAVAYVREGLDVSIPIPSGDVVVVSTDAL 1440
 QY 1441 MTGFTGDESDSDCMTCVTQTVDFLDPTFTTETTTVPQDAVSRQRRGRGSGIYR 1500
 Db 1441 MTGFTGDESDSDCMTCVTQTVDFLDPTFTTETTTVPQDAVSRQRRGRGSGIYR 1500
 QY 1501 FVTPGERPSGMPDSVLCSECYDAGCAWYELTTPAETSVRLRAYLNTFGUVCQDHLFEWES 1560

Db 1501 FVAPGERPSGMPDSVLCSECYDAGCAWYELTTPAETSVRLRAYLNTFGUVCQDHLFEWEG 1560
 QY 1561 VPTGLTHIDAHFLSOTKQAGNFPVLYAVQATVCARAQAPPSPSDQMWKCLIRLKPTLHG 1620
 Db 1561 VPTGLTHIDAHFLSOTKQAGNFPVLYAVQATVCARAQAPPSPSDQMWKCLIRLKPTLHG 1620
 QY 1621 PTPLLYRLGAVONEVILTEPITKYIMACMSADLEVVTWLVGVLAALAAAYCLITCSV 1680
 Db 1621 PTPLLYRLGAVONEVILTEPITKYIMACMSADLEVVTWLVGVLAALAAAYCLITCSV 1680
 QY 1681 VTUGRIILSGPAPVPPDRVLYOSEDMEBECASQIPYIEQGMOLAEQKQKALGLLOAT 1740
 Db 1681 VTUGRIILSGPAPVPPDRVLYOSEDMEBECASQIPYIEQGMOLAEQKQKALGLLOAT 1740
 QY 1741 KOAEAAAPVWESKRALETFWAKHWNFIISGQYLAGLSTLPGNPAIASLMAFTASITSP 1800
 Db 1741 KOAEAAAPVWESKRALETFWAKHWNFIISGQYLAGLSTLPGNPAIASLMAFTASITSP 1800
 QY 1801 LTTQNTLFPNILGWWAAQAPPSAASAPVPGAGIAGAAVSGIGLKVLDIILAGYGAGVA 1860
 Db 1801 LTTQNTLFPNILGWWAAQAPPSAASAPVPGAGIAGAAVSGIGLKVLDIILAGYGAGVA 1860
 QY 1861 GALVAFKWSGVPSTEDLWLLPALSPGALVGVVCAAILRRHVGPGEAVQVWNRLL 1920
 Db 1861 GALVAFKWSGVPSTEDLWLLPALSPGALVGVVCAAILRRHVGPGEAVQVWNRLL 1920
 QY 1921 AFASRGNHVSPTHYHPESDAAARVTQILLSLITQILLKELHQMINEDCSTPCSGSWLRDV 1980
 Db 1921 AFASRGNHVSPTHYHPESDAAARVTQILLSLITQILLKELHQMINEDCSTPCSGSWLRDV 1980
 QY 1981 WDMICVLDFTKTLWOSKLLPRLPGVFPFLSCORGKGVWRGDIQMOTTCPCGAQIAGHVK 2040
 Db 1981 WDMICVLDFTKTLWOSKLLPRLPGVFPFLSCORGKGVWRGDIQMOTTCPCGAQIAGHVK 2040
 QY 2041 NGSMRVLGSPRTCSNTHGTFPINAYTTGCTPSPADNYSRALNRVAABEYVETRVGDEH 2100
 Db 2041 NGSMRVLGSPRTCSNTHGTFPINAYTTGCTPSPADNYSRALNRVAABEYVETRVGDEH 2100
 QY 2101 YVYGMTTDNVKPCQVPAPEFTEVDGVLHRYAPACKPFLLEDVTFQVGLNOYLVGSOL 2160
 Db 2101 YVYGMTTDNVKPCQVPAPEFTEVDGVLHRYAPACKPFLLEDVTFQVGLNOYLVGSOL 2160
 QY 2161 PCSPBPDPVTLTSMITDPSHITAEYAKRRLARQSPSPSLASSASQLSAPSLKATCTHHD 2220
 Db 2161 PCSPBPDPVTLTSMITDPSHITAEYAKRRLARQSPSPSLASSASQLSAPSLKATCTHHD 2220
 QY 2221 SPADLIEANLLWROBMGNITRESEKVKVILDSPEPLHAEGDEREISVAARILEKSRK 2280
 Db 2221 SPADLIEANLLWROBMGNITRESEKVKVILDSPEPLHAEGDEREISVAARILEKSRK 2280
 QY 2281 FPSALPTIWARPDYNPPLLESWKDPDYVVPVYVHGCPLPPTKAPPIPPRRKRTVVLTESNV 2340
 Db 2281 FPSALPTIWARPDYNPPLLESWKDPDYVVPVYVHGCPLPPTKAPPIPPRRKRTVVLTESNV 2340
 QY 2341 SSALAEALATKTRSSGSSAVDSGTATLPALSDDDGDKGSDVBSYSGMPLEGECDPDL 2400
 Db 2341 SSALAEALATKTRSSGSSAVDSGTATLPALSDDDGDKGSDVBSYSGMPLEGECDPDL 2400
 QY 2401 SDGWSVTSVEEA-SEDEVVCCSMSTYTGALITPCAAEESKLPINPLNSLLRHNMVYAT 2459
 Db 2401 SDGWSVTSVEEA-SEDEVVCCSMSTYTGALITPCAAEESKLPINPLNSLLRHNMVYAT 2459
 QY 2460 TSRSASJROKQVTFDRLOVLDHRYDVLKMKKAKASTVYKAKLLSIBEAACKLTPHSAKSK 2519
 Db 2460 TSRSASJROKQVTFDRLOVLDHRYDVLKMKKAKASTVYKAKLLSIBEAACKLTPHSAKSK 2519
 QY 2520 FGYKADVRNLSRRANVHRSWEDILEDTEPTDITIMAKSEVFCVQPKGGRKPARLI 2579
 Db 2520 FGYKADVRNLSRRANVHRSWEDILEDTEPTDITIMAKSEVFCVQPKGGRKPARLI 2579
 QY 2580 FGYKADVRNLSRRANVHRSWEDILEDTEPTDITIMAKSEVFCVQPKGGRKPARLI 2639
 Db 2580 FGYKADVRNLSRRANVHRSWEDILEDTEPTDITIMAKSEVFCVQPKGGRKPARLI 2639

```

Db 2581 VPPDLVRVCEKALYDVVSKLPLAVMGSSYGFQYSPGQVRVPLVQAWKSKKTPMGFSYD 2640
QY 2640 TRCFDSTVTSIRVBSIYQCCDLAPARQATRSITERLYICGPLTNSKGQNCYRRCR 2699
Db 2641 TRCFDSTVTSIRVBSIYQCCDLAPARQATRSITERLYICGPLTNSKGQNCYRRCR 2700
QY 2700 ASGVLTSQNTLTVKATAACRAKLOCTMLVNGDDLWVTCESAGTQEDAAALRAFT 2759
Db 2701 ASGVLTSQNTLTVKATAACRAKLOCTMLVNGDDLWVTCESAGTQEDAAALRAFT 2760
QY 2760 EAMTRYSAPPDPOPEYDELELITSCSNVSVVHADSGKRVYVLTDRPTTPLARAWEA 2819
Db 2761 EAMTRYSAPPDPOPEYDELELITSCSNVSVVHADSGKRVYVLTDRPTTPLARAWEA 2820
QY 2820 RHPFINSWLNIIIMYAPTLWARMILMTHFFSIIILAOBLSKALDCQYIGACYSIEPDLDP 2879
Db 2821 RHPFINSWLNIIIMYAPTLWARMILMTHFFSIIILAOBLSKALDCQYIGACYSIEPDLDP 2880
QY 2880 QITERLHGLSAFTLHYSYSGEINRVASCLRLKLGVPPIETWHRARSVRKILSOGGEAAT 2939
Db 2881 PIQRLHGLSAFTLHYSYSGEINRVASCLRLKLGVPPIETWHRARSVRKILSOGGEAAT 2940
QY 2940 CGYLFNVAVRTKLTPIPAAGOLDLSGMFVAGYSGGDIYHLSLRAPRPFPLCLLLS 2999
Db 2941 CGYLFNVAVRTKLTPIPAAGOLDLSGMFVAGYSGGDIYHLSLRAPRPFPLCLLLS 3000
QY 3000 VGYGYVLLPNR 3010
Db 3001 AGVGILLENR 3011

```

RESULT 8

```

US-09-917-563-2
; Sequence 2, Application US/09917563
; Publication No. US20030073080A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/917,563
; FILING DATE: 27-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/238,076
; FILING DATE: 26-JAN-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3012 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

;
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-917-563-2
Query Match      88.3%; Score 14143.5; DB 10; Length 3012;
Best Local Similarity 85.6%; Pred. No. 0;
Matches 2576; Conservative 218; Mismatches 216; Indels 1; Gaps 1;
QY 1 MSTNPKPQSKTKRNTNRRPQDVKFPGGQIVGGVILLPRRGPRLGVRAIRKASERSQPRG 60
Db 1 MSTNPKPQSKTKRNTNRRPQDVKFPGGQIVGGVILLPRRGPRLGVRAIRKASERSQPRG 60
QY 61 RRQPIPKARRPPEGRAWAQPGYWPVLYGNBGLWAGWLLSPGRSPRWGTTDPRRRSRNLG 120
Db 61 RRQPIPKARRPPEGRAWAQPGYWPVLYGNBGLWAGWLLSPGRSPRWGTTDPRRRSRNLG 120
QY 121 KVIDTITCGFADLMGYPIPLVGAFLGGAARALAHGVVELEDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTITCGFADLMGYPIPLVGAFLGGAARALAHGVVELEDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTIPASAVEVRNVSGIYHVNTDCSNSSIVYEADVIMHTPPGCVPCVQEGNSRCWV 240
Db 181 LLSCLTIPASAVEVRNVSGIYHVNTDCSNSSIVYEADVIMHTPPGCVPCVQEGNSRCWV 240
QY 241 ALTPTLAARNASVPTTIRRHVDLLVGTAAFCGAMTVGDLGSGIFLVSLQFTSPRRHET 300
Db 241 AVTPTVATRDGKLTQLRHRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRRHT 300
QY 301 VQDCNCSIYPGHVSGHRMADMMNNSPTTALVVSOLLRIPOAVDMVAGAHGVLGAGLA 360
Db 301 TQDCNCSIYPGHITGHRMADMMNNSPTTALVVAQLRIPOALMDMIAGAHGVLGAGLA 360
QY 361 YGSMVGNWAKVLIIVALLFAGVDGTHITGTVAGHTTSGFTSLFSSGASQKIQIWNINGSW 420
Db 361 YPSMVGWAKVLIIVALLFAGVDGTHITGTVAGHTTSGFTSLFSSGASQKIQIWNINGSW 420
QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCGPERMASCRPIDWFAQCGWPIYTKENSS 480
Db 421 HINSTALNCNDSINTGWLGLFYCHKFNSSGCGPERLASCRRLTDFAGCGWPIYANGSGL 480
QY 481 DQRPYCHWYAPRCGVVPASQVGVVCTPSPVAVGTDRSGVPTVYSCNENZTDVMLLN 540
Db 481 DERPYCHWYAPRCGVVPASQVGVVCTPSPVAVGTDRSGVPTVYSCNENZTDVFLN 540
QY 541 NTRPPQGNWFGCTWANSSTGFTKTCGGPPCNIGGVNRTLICPTDCRKHPEATYTKCGSG 600
Db 541 NTRPPQGNWFGCTWANSSTGFTKTCGGPPCNIGGVNRTLICPTDCRKHPEATYTKCGSG 600
QY 601 PWLTPCLVDYPYRLMHPYCTLNFSIPKVMYVGGVEHRLNAAACNWTGRGCMLEBDRS 660
Db 601 PWLTPRCMDYDYPYRLMHPYCTINVTIPKVMYVGGVEHRLNAAACNWTGRGCMLEBDRS 660
QY 661 ELSPELLSTTEWQILPCAFITLALSTGLIHLHONIVDVOYLYGVGSAPYSFAIKMEYIL 720
Db 661 ELSPELLSTTQWQVLPCTFTLPAISLGLIHLHONIVDVOYLYGVGSASIAWAKMEYIV 720
QY 721 LLFLLADARVCACLMMMLLIAQAAALENLVLNAAASVAGAGILSFLVFFCAAWIKG 780
Db 721 LLFLLADARVCSCLMMMLLSQAAALENLVLNAAASVAGAGILSFLVFFCAAWIKG 780
QY 781 RLAPCAAYAFYGVWPLLLLLALPPRAYALDREVAASGGGAVLVGLVFLTLSPYKVFELT 840
Db 781 RWFGVAVAFYGMWPLLLLLALPPRAYALDREVAASGGGAVLVGLVFLTLSPYKVFELT 840
QY 841 RLIIWVLYFITEAAHQVWVPPVPLNVEGGDAITLITCAVHPELIFDITKLLIALLCPML 900
Db 841 WCMWVLYFITEAAHQVWVPPVPLNVEGGDAITLITCAVHPELIFDITKLLIALLCPML 900
QY 901 VLOAGITRVFVRAQGLIRACMLVRKVAGGHYVQVPMKLGALITGYVYNHLLTFLDWA 960
Db 901 ILQASLLKVPFVRAQGLIRACMLVRKVAGGHYVQVPMKLGALITGYVYNHLLTFLDWA 960

```

```

Qy 961 HAGLRDLAVAVEPVVFSAMETKVTITWADTAACGDIILGLPVARRGKEIFLGPADSLBG 1020
Db 961 HNGLRDLAVAVEPVVFSAMETKLTWADTAACGDIILGLPVARRGKEILLGPADGMSV 1020
Qy 1021 QGWELLAPITAYSQOTRGVLCIIITSTIGRDKNOVEGEVQVSTATOSFIATCINGVCWT 1080
Db 1021 KGWELLAPITAYAOQTRGLLCIIITSTIGRDKNOVEGEVQVSTATQTFIATCINGVCWT 1080
Qy 1081 VYHAGSKTLGAPGPITOMYNTVNDLIVGWQAAPPGARSTPCSCGSSDLYLVTRHADVI 1140
Db 1081 VYHAGSKTLGAPGPITOMYNTVNDLIVGWQAAPPGARSTPCSCGSSDLYLVTRHADVI 1140
Qy 1141 PVRRGDSRSGLLSPRPVSYLKGSGGPFLLCPSPGHVGVGFRAAYCTRGVAKAVDFIPVES 1200
Db 1141 PVRRGDSRSGLLSPRPVSYLKGSGGPFLLCPSPGHVGVGFRAAYCTRGVAKAVDFIPVEN 1200
Qy 1201 METWRSFVTDNSTTPRAVPTQVVAHLHAPTGSKSTKVPAAVAAOQYKVLVNPVAA 1260
Db 1201 LETWRSFVTDNSTTPRAVPTQVVAHLHAPTGSKSTKVPAAVAAOQYKVLVNPVAA 1260
Qy 1261 TLGFGAYMSKAHGIDPNIRTCVRTITTTGGSITYSTYKFLADGGCGGAYDIIICDECHS 1320
Db 1261 TLGFGAYMSKAHGIDPNIRTCVRTITTTGGSITYSTYKFLADGGCGGAYDIIICDECHS 1320
Qy 1321 TDSTILGIGTVLDOAETAGARLVVLATATPPGSVTVFHPNIEHIGLSNNGEIPFYGKAI 1380
Db 1321 TDSTILGIGTVLDOAETAGARLVVLATATPPGSVTVFHPNIEHIGLSNNGEIPFYGKAI 1380
Qy 1381 PIEAIKGRHILFCHSKKKCBELAAKULCLGINAVAYYRGLDVSVIPPIGVDVVVATDAL 1440
Db 1381 PLEVIKGRHILFCHSKKKCBELAAKULCLGINAVAYYRGLDVSVIPTSGDVVVVATDAL 1440
Qy 1441 MTGFTGDFSDVDNCTVQVDFSLDPTFTTETTVQDVAVSRRGRGRGRSGIYR 1500
Db 1441 MTGFTGDFSDVDNCTVQVDFSLDPTFTTETTVQDVAVSRRGRGRGRSGIYR 1500
Qy 1501 FVTPGERPSGMFDDSVLCEYDAGCAWYELTPAETSVELRAYLNTPLPVPQDHLRFWES 1560
Db 1501 FVAPGERPSGMFDDSVLCEYDAGCAWYELTPAETTVELRAYLNTPLPVPQDHLRFWEG 1560
Qy 1561 VFTGLTHIDAHFLSOTKAGDNFPYLVAQATVCARAQAPPPSDOMWKCLIRLKPETHG 1620
Db 1561 VFTGLTHIDAHFLSOTKAGDNFPYLVAQATVCARAQAPPPSDOMWKCLIRLKPETHG 1620
Qy 1621 PTPLLYRLGAVONEVILTHPTKYTMAQMSADLEVVTVTLVGVGLAALAAAYCITTTGSV 1680
Db 1621 PTPLLYRLGAVONEVILTHPTKYTMAQMSADLEVVTVTLVGVGLAALAAAYCITTTGSV 1680
Qy 1681 VIVGRILLSGKPAVVDREVLYQRPDEMEBCASQOLPIEQGQIAEQKQKALGLLOAT 1740
Db 1681 VIVGRILLSGKPAVVDREVLYQRPDEMEBCASQOLPIEQGQIAEQKQKALGLLOAT 1740
Qy 1741 KOAEAAAAPVBSKRALETFWAKHWNFTSGIYQYLAGLSTLPGNEPAIASLMAFTASITSP 1800
Db 1741 ROAEVITPAVQVNTWOKLEVFWAKHWNFTSGIYQYLAGLSTLPGNEPAIASLMAFTASITSP 1800
Qy 1801 LITQNTLIFNLIGGWAAQAPPSAASAFVAGIAGAAGVSGTGLGKVLVDIILAGYGAGVA 1860
Db 1801 LITQNTLIFNLIGGWAAQAPPSAASAFVAGIAGAAGVSGTGLGKVLVDIILAGYGAGVA 1860
Qy 1861 GALVAFKMGSGEVPSTEDJVNLLPAILSPGALVGVVCAAILRRHVGPGEVQVWNNRLI 1920
Db 1861 GALVAFKMGSGEVPSTEDJVNLLPAILSPGALVGVVCAAILRRHVGPGEVQVWNNRLI 1920
Qy 1921 AFASRGNHVSPTHYVPESDAAAEVTOILSSLTITOLKELHWINEDCQPCSGSWLRDV 1980
Db 1921 AFASRGNHVSPTHYVPESDAAAEVTOILSSLTITOLKELHWINEDCQPCSGSWLRDI 1980
Qy 1981 WDWICVTLDFTKWLQSKLLPRLPVPFFLSQKGYKGVWREGDITMOTTCPOGAQIAGHYK 2040
Db 1981 WDWICEVLSDFKTLWAKALMPQLPGIFVFCQKGYKGVWREGDITMOTTCPOGAQIAGHYK 2040
Qy 2041 NGSMRIUGPRTCSNTHGTFPPINAYTTGCTPSPAPNYSGALWRVLAEEYVEVTRVGDFH 2100

```

RESULT 9

US-10-296-734-406

; Sequence 406, Application US/10296734

; Publication No. US20040054:37A1

; GENERAL INFORMATION:

```

Db 2041 NGTMRIUGPRTCRNMGSGTFPINAYTTGCTPPLPAPNYKFALMRVSAEBYVIRRVGDFH 2100
Qy 2101 YVTGMTTDNVKPCQVPAPEFFTEVDGVLHAYAPACKPLLREDVTFQVGLNOYLVGSQL 2160
Db 2101 YVTGMTTDNVKPCQVPAPEFFTEVDGVLHAYAPACKPLLREDVTFQVGLNOYLVGSQL 2160
Qy 2161 PCPEPDDVTVLTSMLTDPHSITAEAKRRLARGSPESLASASSASQLSAPSLKATCTTHHD 2220
Db 2161 PCPEPDDVTVLTSMLTDPHSITAEAKRRLARGSPESLASASSASQLSAPSLKATCTTHHD 2220
Qy 2221 SPDAELIEANLLWRQMGNNITRVESENKVVILDSFSP2LHAECDEREISVAABILLKSRK 2280
Db 2221 SPDAELIEANLLWRQMGNNITRVESENKVVILDSFSP2LHAECDEREISVAABILLKSRK 2280
Qy 2281 FPSALPIWAPDPYNNPILLESWKDPDYVPVHVHGCPLPPTKAPP2IPPRRRKRTVVLTESNV 2340
Db 2281 PARALPWAPDPYNNPILLESWKDPDYVPVHVHGCPLPPTKAPP2IPPRRRKRTVVLTESNV 2340
Qy 2341 SSALAEATKTFGSSGSSAVDSOTATALPDLASDDGDKGSDVESYSSMPPLEGEPPDL 2400
Db 2341 STALAEATKTFGSSGSSAVDSOTATALPDLASDDGDKGSDVESYSSMPPLEGEPPDL 2400
Qy 2401 SDGWSVTVSEEA-SEOVVCCMSVYTWGALITPCAABESKLPINPLSNSLRHNNMVYAT 2459
Db 2401 SDGWSVTVSEEA-SEOVVCCMSVYTWGALITPCAABESKLPINPLSNSLRHNNMVYAT 2459
Qy 2460 TSSASLRQKQVTFDRLOVLDHVDLKEMKAKASTVKAKLISIEEACKLTTPHSAKS 2519
Db 2460 TSSASLRQKQVTFDRLOVLDHVDLKEMKAKASTVKAKLISIEEACKLTTPHSAKS 2519
Qy 2520 FGYGAKDVNLSRANVHRSWEDLLETTETPIDITTIMAKSEVFCVQPKGGRKPARLI 2579
Db 2520 FGYGAKDVNLSRANVHRSWEDLLETTETPIDITTIMAKSEVFCVQPKGGRKPARLI 2579
Qy 2580 VFPDLGVRCVKMALYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTWKCKCPMGFSYD 2639
Db 2580 VFPDLGVRCVKMALYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTWKCKCPMGFSYD 2639
Qy 2640 TRCFDSVTBESDIRVEESIYCCCDLAPAPARQAIRSLTERLYIGGPLTNSKGONCYRRCR 2699
Db 2640 TRCFDSVTBESDIRVEESIYCCCDLAPAPARQAIRSLTERLYIGGPLTNSKGONCYRRCR 2699
Qy 2700 ASGVLTTSCGNLTFCYKATACAAKRAKLOCTMLVNGDGLWICESAGTQEDAAALBAFT 2759
Db 2700 ASGVLTTSCGNLTFCYKATACAAKRAKLOCTMLVNGDGLWICESAGTQEDAAALBAFT 2759
Qy 2760 EAMTRYGAPDPPOPEVDLELITSCSSNVSAHDAGKRVYVLTDRPTTFLAAZAAETA 2819
Db 2760 EAMTRYGAPDPPOPEVDLELITSCSSNVSAHDAGKRVYVLTDRPTTFLAAZAAETA 2819
Qy 2820 RHTPINSWLGNIIMYAPTLWARMILMTFFSILLABQBLEKALDCQYGYACYSIEPLDLP 2879
Db 2820 RHTPINSWLGNIIMYAPTLWARMILMTFFSILLABQBLEKALDCQYGYACYSIEPLDLP 2879
Qy 2880 QIIERLHGLSFTLHSPGEBINRVAACLRKLGVPPLTWHRARSVRKLLSOGGAAT 2939
Db 2880 QIIERLHGLSFTLHSPGEBINRVAACLRKLGVPPLTWHRARSVRKLLSOGGAAT 2939
Qy 2940 CGRYLFNNAVATKLTLP2PAASOLD-SGWFWAGVSGGDIYVESLSRARPRPWFPLCLLLS 2999
Db 2940 CGRYLFNNAVATKLTLP2PAASOLD-SGWFWAGVSGGDIYVESLSRARPRPWFPLCLLLS 2999
Qy 3000 VGVGIYLLPNR 3010
Db 3001 AGVGIYLLPNR 3011

```

; APPLICANT: Thompson, Scott A
 ; APPLICANT: Ramshaw, Ian A
 ; TITLE OF INVENTION: Synthetic molecules and uses therefor
 ; FILE REFERENCE: Savine
 ; CURRENT APPLICATION NUMBER: US/10/296,734
 ; CURRENT FILING DATE: 2003-08-04
 ; PRIOR APPLICATION NUMBER: AU P07761/00
 ; PRIOR FILING DATE: 2000-05-26
 ; NUMBER OF SEQ ID NOS: 1507
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 406
 ; LENGTH: 3011
 ; TYPE: PR
 ; ORGANISM: Artificial
 ; FEATURE:
 ; OTHER INFORMATION: HepC la consensus polyprotein
 US-10-296-734-406

Query Match		88.3%;	Score 14131.5;	DB 12;	Length 3011;
Best Local Similarity		85.5%;	Pred. No. 0;		
Matches 2574;		Conservative 216;	Mismatches 220;	Indels 1;	Gaps 1;
Qy	1	MSNPKPQRTKRNTRRRPDDVPFGGQIVGGVYLPRRQPLRGVATRKASERSQPRG	60		
Db	1	MSNPKPQRTKRNTRRRPDDVPFGGQIVGGVYLPRRQPLRGVATRKASERSQPRG	60		
Qy	61	RRQPIKARREGEAWAQPYPWLYNGEGLGWAGWLLSPRGSPSPGTPDPRRSRNLG	120		
Db	61	RRQPIKARREGEAWAQPYPWLYNGEGLGWAGWLLSPRGSPSPGTPDPRRSRNLG	120		
Qy	121	KVIDTLGCGPADLMGYIPLVGAPLGGARALAHGVRLVDGNYATGNLPGCSFSIFLLA	180		
Db	121	KVIDTLGCGPADLMGYIPLVGAPLGGARALAHGVRLVDGNYATGNLPGCSFSIFLLA	180		
Qy	181	LLSCLITPASAYEVNRVSGIYHVTNDCNSIIVAEADVIHMTGCVPCVCGNSRCRW	240		
Db	181	LLSCLITPASAYEVNRVSGIYHVTNDCNSIIVAEADVIHMTGCVPCVCGNSRCRW	240		
Qy	241	ALTPTLAARNASVPTTTRRVDVLLVGTAAFCSAMVGDLCGSIPLVSOQLTFFPRHET	300		
Db	241	AMTPTVATRDGKLPAIQRRHIDLLVGSATLCSALYVGDLCGSIPLVSOQLTFFPRHET	300		
Qy	301	VQDNCNSIYPGHVSGHRMAMDMNWSPTTALVWSQLLRIPQAVVDMVAGAHGWLAGLA	360		
Db	301	TQDNCNSIYPGHVSGHRMAMDMNWSPTTALVWSQLLRIPQAVVDMVAGAHGWLAGLA	360		
Qy	361	YYSVMGNWAKVLIYVALLPAGVDGETHTRVAGHTTSFTSLFSSGASQKQLVNTNGSW	420		
Db	361	YFSVMGNWAKVLIYVALLPAGVDGETHVTCGNAGRTTSLGLVSLITPGAKQNTQLINTNGSW	420		
Qy	421	HINFTALNCNDSLOQGFPAALFYAHKENSNGCPRMASCRPIDWFAQGWGPTTYTKPNS	480		
Db	421	HINSTALNCNDSLNTGWLHAGFYQHKFNSSGCPERLASCRRLTDFDQNGPISYANGSGP	480		
Qy	481	DQPYCMHYAPRCPGWVSPASQVCGPVYCTPSPVVVGTDRSGVPTYSNGENETDNLN	540		
Db	481	DQPYCMHYAPRCPGIVPAKSVCGPVYCTPSPVVVGTDRSGVPTYSNGANDTDVFLN	540		
Qy	541	NTRPPQGNWFGCTWMNSTGFTTCGPPQNIIGVGNRTLIPTDCFRKHPEATYKCGSG	600		
Db	541	NTRPPLGNWFGCTWMNSTGFTTCGAPPVIGAGNNTLHCTPDCFRKHPEATYKCGSG	600		
Qy	601	PMLTRCLVDYPRVLMWHYPCNTLNFSEFKVMVGVGVEHRLNACNWRGRCNLEDRDRS	660		
Db	601	PMITRCLVDYPRVLMWHYPCNTNYTLFKVMVGVGVEHRLNACNWRGRCNLEDRDRS	660		
Qy	661	ELSPILLSTTEWQILPCAPTTLPALSTGLIHLFONIVDVQYIYGVGSAPVSPFAIKWBYLL	720		
Db	661	ELSPILLSTTQCVLPSCFTTLPALSTGLIHLFONIVDVQYIYGVGSIIASNAIKWEYV	720		
Qy	721	LLFLLIADARVCSCLMMLLIQAQAALNLLVNLNAAVAGAHGIIISFLVFFCAAWYIKG	780		
Db	721	LLFLLIADARVCSCLMMLLIQAQAALNLLVNLNAAVAGAHGIIISFLVFFCAWYIKG	780		

Qy	781	RLAPGAAYAFYGVWPLLLLLALLPPRAYALDREMAASCGGAVLVGLVFTLTSPYKVFILT	840		
Db	781	RWVPGAVYALYGVWPLLLLLALLPQRAYALDTEVAASCGGVVAVGLMALLTSPYKRYTS	840		
Qy	841	RLIWLQYBITRAEAHMQWVPLNVGRGSDAILLTCAVHPELFDITKLLAILLGLPLM	900		
Db	841	WCLMWLQYFLTRVEAQLHVVPPLNVGRGRDAVILLVCVVHFTLVFDITKLLAVFGPLM	900		
Qy	901	VLOAGITRVPIYFVRAOGLIRACMLVKVAGGHVYQVFMKLGALTGTYVYNHLTPLRDWA	960		
Db	901	ILQASLLKVPYFVRVQGLLRIICALARKMLGGHVQNAIILKLGALTGTYVYNHLTPLRDWA	960		
Qy	961	HAGLRDLAVAVEPVPFSAMETKVITWGAOTACGDIILGLPVSARAKGIFLGPADSLSE	1020		
Db	961	HNGLRDLAVAVEPVPFSQMETKLIITWGAOTACGDIINGLPVSARRGREILLGPADQWVS	1020		
Qy	1021	QGWLLAPITAYSQOTRGVLGCIITSLTGEDKNQVEGEVQVYSTATQSLFATCINGVCWT	1080		
Db	1021	KGWLLAPITAYAAQQTGRLGCIITSLTGRDRKNQVEGEVQIVSTAAQTFLATCINGVCWT	1080		
Qy	1081	VYHGAGSKTLAGEKGPITOMYTNVDLDLVQWQAPPGARSMTPCSCGSSDLYLVTRHADVI	1140		
Db	1081	VYHGAGTRIITASPKGPVIQMYTNVDQDLVQWPAQSGRSRLTPTCTCGSSDLYLVTRHADVI	1140		
Qy	1141	PVARRGDSRGLSLGPRPVSYLKSSGGPLLCPSGHVGVPPRAAVCTRGVAKAVDPTPVS	1200		
Db	1141	PVARRGDSRGLSLSPRPISYLGKSSGGPLLCPAAGHAVGIFRAAVCTRGVAKAVDPTFVEN	1200		
Qy	1201	METMRSPTVDKSTPPAVPQTFQVAHLHAPTSGSKTKVPAAYAAQGYKVLVNPVAA	1260		
Db	1201	LETMRSPTVFTDNSSPPAVPQSFQVAHLHAPTSGSKTKVPAAYAAQGYKVLVNPVAA	1260		
Qy	1261	TLGFGYMSKXAGIDPNIRTVRTITTTGGSTITYSTYKFLADGSCGAGYDIIICDSCHS	1320		
Db	1261	TLGFGYMSKXAGIDPNIRTVRTITTTGGSPITYSTYKFLADGSCGAGYDIIICDSCHS	1320		
Qy	1321	TDSTTLIGITVLDQAEATAGARLVLAATAIPPGQSVTPHPNIEIIGLNNGEIPPYCKAI	1380		
Db	1321	TDATSLIGITVLDQAEATAGARLVLAATAIPPGQSVTPHPNIEEVALSTTGEIPPYCKAI	1380		
Qy	1381	PIEAKGRHLIFCHSKKCCDELAALGTGCLNNAVAVRGVLDVSVIPIIDGVVVVATDAL	1440		
Db	1381	PLEVIGKGRHLIFCHSKKCCDELAALVALGINNAVAYRGVLDVSVIPTSGLVVVATDAL	1440		
Qy	1441	MTGFTGDFSDVIDCNTCVTQVDFSLDPTFTIETTTVPQDAVSQRSQRTGRGSGIYR	1500		
Db	1441	MTGYTGDFSDVIDCNTCVTQVDFSLDPTFTIETTTLPQDAVSRQRTGRGSGIYR	1500		
Qy	1501	PVTGGERPSGMFDSVLCBCYDAGCAWYELTPAETSURLRAYLNTRGHPYCODHLEFWES	1560		
Db	1501	FVAPGERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTFGLPVCQDHLFEWEG	1560		
Qy	1561	VFTGLTHIDAHFLSOTQAGDNFPYLVAYQATVCARAQAPPPSWDQWKKLIRLKPTLHG	1620		
Db	1561	VFTGLTHIDAHFLSOTQSGENFPYLVAYQATVCARAQAPPPSWDQWKKLIRLKPTLHG	1620		
Qy	1621	PTPLLYRLGAVQNEVILTHPTIKYIMACMSADLEBVVTSWTVLVGVGLAALAAAYCLTTGSV	1680		
Db	1621	PTPLLYRLGAVQNEVILTHPTKIMTMSADLEBVVTSWTVLVGVGLAALAAAYCLSTGCV	1680		
Qy	1681	VIVGRILISGKPAVPPREVLYQBFDEMEBCASQIPYEQGMOLAEQPKQKALGLIQTAT	1740		
Db	1681	VIVGRIVLSGKPAIIPREVLYRBFDEMEBCSQHLPYIEQGMILARQPKQKALGLIQTAS	1740		
Qy	1741	KQAEAAAPVVSRYKPALETTFWAKMWNFIISGYOVLAGLSTLPGNPAIASLMAFTASITSP	1800		
Db	1741	ROAEVIAPVOTVNTWOKLEVFWAKENWNFIISGYOVLAGLSTLPGNPAIASLMAFTAVTSP	1800		
Qy	1801	LTTQNTLLFNILGCVAAQAALAPPASAASAFVAGAGIAGAAVSGISGLKVLIVDLIAGYGAV	1860		
Db	1801	LTTSQTLFNILGCVAAQAALAPGATATFVAGLAGAAGIAGVGLKVLIVDLIAGYGAV	1860		

QY 1861 GALVAFKUMSGVSEPTEDLVNLLPAILSPGALVGVVCAAILRRHHVGPCEGAVQVMMRLI 1920
 Db 1861 GALVAFKUMSGVSEPTEDLVNLLPAILSPGALVGVVCAAILRRHHVGPCEGAVQVMMRLI 1920
 QY 1921 AFASRGNHVSPTHYVPESDAAAEVTOILSLTITQLLKHOMINDECSPTCSGSLWLDV 1980
 Db 1921 AFASRGNHVSPTHYVPESDAAAEVTOILSLTITQLLKHOMINDECSPTCSGSLWLDV 1980
 QY 1981 KDWICVLTDFKTLWLOSKILLPRLPGVPEFLSCQGYKGVWRGDIOMOTPCGQAOIAGHVK 2040
 Db 1981 KDWICVLTDFKTLWLOSKILLPRLPGVPEFLSCQGYKGVWRGDIOMOTPCGQAOIAGHVK 2040
 QY 2041 NGSMRIVGRTCNTHGTHPPINAYITGCTPSPAPNYSRALMRVAEEYVEVTRYGDFH 2100
 Db 2041 NGSMRIVGRTCNTHGTHPPINAYITGCTPSPAPNYSRALMRVAEEYVEVTRYGDFH 2100
 QY 2101 YVTGMTDNNVKCPQVAPBEPFTEVDGVRHLRYAPACKPLLRDVTFOVGLNOYVLSQL 2160
 Db 2101 YVTGMTDNNVKCPQVAPBEPFTEVDGVRHLRYAPACKPLLRDVTFOVGLNOYVLSQL 2160
 QY 2161 PCSEPEPDTVLTLSMLTDPSTHIAETAKRRLARGSPPSLASSASQSLAPSLKATCTHHD 2220
 Db 2161 PCSEPEPDTVLTLSMLTDPSTHIAETAKRRLARGSPPSLASSASQSLAPSLKATCTHHD 2220
 QY 2221 SPDAELIEANLLWRQEMGNITRVESNKVILDSPEPLHAEGDERISVAABILKRSK 2280
 Db 2221 SPDAELIEANLLWRQEMGNITRVESNKVILDSPEPLHAEGDERISVAABILKRSK 2280
 QY 2281 FPSALPFWARPDVNPPLLESKWDPDYVPPVHGCPLPPTKAPPIPPRRRKTCTVVLDES 2340
 Db 2281 FPSALPFWARPDVNPPLLESKWDPDYVPPVHGCPLPPTKAPPIPPRRRKTCTVVLDES 2340
 QY 2341 SSALAEALATKTFGSGSSAVDSGTATLPDLASDDGKGDVSSYSSMPPLEGEPPDPL 2400
 Db 2341 SSALAEALATKTFGSGSSAVDSGTATLPDLASDDGKGDVSSYSSMPPLEGEPPDPL 2400
 QY 2401 SDGSWSTVSPEA-SEDVVCSSMSTWTGALITPOAAESKLPINPLNSLLRHNNMVIAT 2459
 Db 2401 SDGSWSTVSPEA-SEDVVCSSMSTWTGALITPOAAESKLPINPLNSLLRHNNMVIAT 2459
 QY 2460 TSRSASTROKVTDRLOVLDHVEDVLKEMKAASVKAKLISIEACKLTTPESAKSK 2519
 Db 2460 TSRSASTROKVTDRLOVLDHVEDVLKEMKAASVKAKLISIEACKLTTPESAKSK 2519
 QY 2520 FGYGAKDVRNLSRAVNHRSVMDLLEDTEPTDITIMAKSEVFCVQPKGGRKPARLI 2579
 Db 2520 FGYGAKDVRNLSRAVNHRSVMDLLEDTEPTDITIMAKSEVFCVQPKGGRKPARLI 2579
 QY 2580 VFPDLGVRVCKMALYDVVSTLPOAVMGSSYGFQYSPQRFVFLVNTWKSKKCPMGFSYD 2639
 Db 2580 VFPDLGVRVCKMALYDVVSTLPOAVMGSSYGFQYSPQRFVFLVNTWKSKKCPMGFSYD 2639
 QY 2640 TRCFDSTVTESDIRVESIYOCCLAPARQAIISLTERLYIGGLTNSKQNGCYRRCR 2699
 Db 2640 TRCFDSTVTESDIRVESIYOCCLAPARQAIISLTERLYIGGLTNSKQNGCYRRCR 2699
 QY 2700 ASGVLTTCSCGNTLCYLKATAACRAAKLQDCTMLVNGDDLVIIVICESAGTOEDAAALRAFT 2759
 Db 2700 ASGVLTTCSCGNTLCYLKATAACRAAKLQDCTMLVNGDDLVIIVICESAGTOEDAAALRAFT 2759
 QY 2760 BAMTRYGAPPGDPPOPEVDLELITSCSSNVSAHDASGRVYVLTTRDPTPLARAANETA 2819
 Db 2760 BAMTRYGAPPGDPPOPEVDLELITSCSSNVSAHDASGRVYVLTTRDPTPLARAANETA 2819
 QY 2820 RHTPNSWGLNITVAPPTWARMILMTHTFSTILIAQEOLEKALDCOIVGACVSTEPDLP 2879
 Db 2820 RHTPNSWGLNITVAPPTWARMILMTHTFSTILIAQEOLEKALDCOIVGACVSTEPDLP 2879
 QY 2880 QIERLHGLSAPTHLSYSGFEINRVASCLRLKGVPPPLTWRRHARSVRKALLSOGGRAAT 2939
 Db 2880 QIERLHGLSAPTHLSYSGFEINRVASCLRLKGVPPPLTWRRHARSVRKALLSOGGRAAT 2939
 QY 2940 CORYLFNWAVRTKLTLPIDPAASQDLDSGWFVAGYSGGDIYHSLSRAPRPFPLCLLLLS 2999

Db 2941 CGKYLFWAVATKLTLPIDPAASQDLDSGWFVAGYSGGDIYHSLSRAPRPFPLCLLLLS 3000
 QY 3000 VGVGYLLPNR 3010
 Db 3001 AGVGYLLPNR 3011
 RESULT 10
 US-09-952-572-9
 ; Sequence 9, Application US/09952572
 ; Patent No. US20020119495A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
 ; APPLICANT: NAKANO, Eileen
 ; APPLICANT: CLEMENTS, David
 ; APPLICANT: HUMPHREYS, Tom
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
 ; FILE REFERENCE: HAMBIO1100
 ; CURRENT APPLICATION NUMBER: US/09/952,572
 ; PRIOR FILING DATE: 2001-09-13
 ; PRIOR APPLICATION NUMBER: US 60/230,927
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 9
 ; LENGTH: 3011
 ; TYPE: PRN
 ; ORGANISM: Hepatitis C Virus
 US-09-952-572-9
 Query Match 88.2%; Score 14124.5; DB 9; Length 3011;
 Best Local Similarity 85.5%; Pred. No. 0;
 Matches 2573; Conservative 218; Mismatches 219; Indels 1; Gaps 1;
 QY 1 MSTNPKPQKTKTENTNRRPQDVKFGGGOIVGGVLLPRRGLGVATRKASERSQREG 60
 Db 1 MSTNPKPQKTKTENTNRRPQDVKFGGGOIVGGVLLPRRGLGVATRKASERSQREG 60
 QY 61 RRQPIKARRPREGRAWAFCYWPVLYNGLWAGWLLSPRGSFWSWGTDPRRRSRNLG 120
 Db 61 RRQPIKARRPREGRAWAFCYWPVLYNGLWAGWLLSPRGSFWSWGTDPRRRSRNLG 120
 QY 121 KVTDITTCGADLMGYIPLVGAFLGAAALAHGVLEEDGVNATCNLPGCSFSIFLLA 180
 Db 121 KVTDITTCGADLMGYIPLVGAFLGAAALAHGVLEEDGVNATCNLPGCSFSIFLLA 180
 QY 181 LLSCLTIPASAYEVNRVSGIYHVNTDCSNSSIVYEADVIMHTPPGCVPCVQEGNSRCWV 240
 Db 181 LLSCLTIPASAYEVNRVSGIYHVNTDCSNSSIVYEADVIMHTPPGCVPCVQEGNSRCWV 240
 QY 241 ALPTTLAARNASVPTTIRRHVDLLVGTAAFGSAMVVGDLCSIFLVSQLFTFSPRRHET 300
 Db 241 AVTFTVATRDGKLTTLQLRHHIDLLVGSATLCSALYVGDLCGSFVLVGOLFPTSPRRHWT 300
 QY 301 VQDNCNSIYPGHVSGHMDMMNNSPTALVVSQLLRTPQAVVDMVAGAHGVLGAGLA 360
 Db 301 TQDNCNSIYPGHVSGHMDMMNNSPTALVVSQLLRTPQAVVDMVAGAHGVLGAGLA 360
 QY 361 YSMVGNWAKVLIALLFAGVDGETHTTKRVAHGHTTSGFTSLFSSGASQKIQJNTNGSW 420
 Db 361 YSMVGNWAKVLIALLFAGVDGETHTTKRVAHGHTTSGFTSLFSSGASQKIQJNTNGSW 420
 QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGGERMASCRPIOWPAGWGPIITYKPNSS 480
 Db 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGGERMASCRPIOWPAGWGPIITYKPNSS 480
 QY 481 DQRPYCHWYAPRCGVVYPAQVGVVCTPSPVAVGTDRSGVTVYVSWGENTDVMXLN 540
 Db 481 DQRPYCHWYAPRCGVVYPAQVGVVCTPSPVAVGTDRSGVTVYVSWGENTDVMXLN 540
 QY 541 NTRPPQCNWPGCTWNNSTGFTKTCGGPPCNIGGVNRTLICPTDCFKHPEATYTKCGSG 600

541 NTRPPLGNWFGTWNSTGFTKVCAGAPPCVIGVGNNTLLCPTDCFRKHPEATVSRCSG 600
 601 PWLTPCLVDYDYPRLKHYPCTLNFSIFKRMVYGGVHEHRLNAACNTRGERCNLEDRDS 660
 601 PWITPRCMVDYDYPRLWHYPCTINYTFKRMVYGGVHEHLEAACNTRGERCNLEDRDS 660
 661 ELSPLLSSTTEWOLPCAFITLPALESTGLHHLHONIVDVOYLXGVSFAFSFAIKWEYIL 720
 661 ELSPLLSSTQWQVLPSCFTTLPALESTGLHHLHONIVDVOYLXGVSSTASWALKWEYIV 720
 721 LLELLADARVCACLMMMLLIAQAEALENLVLNAAASVAGAFGLISFLVFFCAANWIKG 780
 721 LLELLADARVCSCLMMMLLISQAERALENLVLNAAASLAGEGLVSLVFFCFAMWLK 780
 781 RLAPGAAYAYGWPLLLLLLLALPPRAYALDREWAACGAVLGLVFLTLSPYKVFLLT 840
 781 RWPFGAVIALYGMWPLLLLLLLALPPRAYALDREWAACGAVLGLVFLTLSPYKRYIS 840
 841 RLJWMLQYFITRABAHQVMWPLNVNVRGGDAIILLTCAVHEPILIFITIKLALLGLPLM 900
 841 WCMWNLQYFITRVEAQHVGWVPLNVNVRGGDAVILLMCVHEPILVFDITKLLALLGLPLW 900
 901 VLQAGITRVPYVRAQGLIRACMLVRKAGHYVQWVMFKLGTGYVYVNHLPLEADWA 960
 901 ILQASLLKVYFVRVQGLLRICALARKIAGHYVQWMAIILKLGALTGYVYVNHLPLEADWA 960
 961 HAGLRDLAVAVEPVPSAMETHKITWGDATACGDIILGLPVSAARRKEIFLGPADSLG 1020
 961 HNGLRDLAVAVEPVPSAMETHKITWGDATACGDIILGLPVSAARRKEIFLGPADGWS 1020
 1021 QGWELLAPIYAYSQOTRGVLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCINGVCWT 1080
 1021 KGWELLAPIYAYAQTRGLLGCIIITSLTGRDKNOVEGEVQVSTATQTFLATCINGVCWT 1080
 1081 VYHAGSKTILAGPKGPTQMTNVDLLVGHQAPPGASMTPCSCGSDLYLVTRHADVI 1140
 1081 VYHAGTRITIASPKGPVQMTNVDQDLVGPAPQGSRLTPCTCGSSDLYLVTRHADVI 1140
 1141 PVRRGDSRGLLSRPSVYLKSGSGPLCPSGHVGVPRAAVCTRGVAKAVDPFIVES 1200
 1141 PVRRGDSRGLLSRPSVYLKSGSGPLCPSGHVGVPRAAVCTRGVAKAVDPFIVEN 1200
 1201 METWRSVPVTDNTPPAVPOTFVAHLHAPTGSGKSTKVPAAYAAQGYKVLNLPVAA 1260
 1201 LGTTWRSVPVTDNTPPAVPOTFVAHLHAPTGSGKSTKVPAAYAAQGYKVLNLPVAA 1260
 1261 TLGFGAYMSKAHGIDPNIRTOVRTITGSGSTYTYGKFLADGGCGGAYDIIICDECHS 1320
 1261 TLGFGAYMSKAHGIDPNIRTOVRTITGSGSTYTYGKFLADGGCGGAYDIIICDECHS 1320
 1321 TDSITLIGTVLDOAETAGARLVVLATATPPGSVTVPHPNIERIGLNGGEIIPYKAI 1380
 1321 TDAITLIGTVLDOAETAGARLVVLATATPPGSVTVPHPNIERIGLNGGEIIPYKAI 1380
 1381 PIRAIKGRHILFCHSKKKDELAAKTGLGINAVAYYRGDLVSIPPIDGVVVVATDAL 1440
 1381 PLEVIKGRHILFCHSKKKDELAAKVALGINAVAYYRGDLVSIPPISGDDVVVSTDAL 1440
 1441 MTGFTGDFSDVIDNCVTQVDFSLDPTFIETITVPODAVRSORGRGSGIYR 1500
 1441 MTGFTGDFSDVIDNCVTQVDFSLDPTFIETITVPODAVRSORGRGSGIYR 1500
 1501 FVTPGERSGFSDSVLCECYDAGCAWYELPBAETSRLRAYLNPFGLPVQODHLEFWES 1560
 1501 FVAPGERSGFSDSVLCECYDAGCAWYELPBAETVRLRAYLNPFGLPVQODHLEFWEG 1560
 1561 VFTGLTHIDAHPLSOTKQAGDNFFYINAVQATVCARAQAPPSWDMKCLRLKPTLHG 1620
 1561 VFTGLTHIDAHPLSOTKQSGENFPYLVAVQATVCARAQAPPSWDMKCLRLKPTLHG 1620
 1621 PTPILYRIGAVONEVILTHPTIKYIMACMSADLEVVSTWLVGGVLAALAYCLITGSV 1680
 1621 PTPILYRIGAVONEVILTHPTIKYIMTCMSADLEVVSTWLVGGVLAALAYCLITGSV 1680

1681 VIVGRILLSGKDAVVPDREVLYQBFDEMEECASQLPYIEQGMOLAEQFKQKALGLQAT 1740
 1681 VIVGRILLSGKDAVVPDREVLYQBFDEMEECASQLPYIEQGMOLAEQFKQKALGLQAT 1740
 1741 KOREAAAAPVBSKRALETFWAKHWNFTISGTOYLAGLSTLPGNPAIASMAFTASITSP 1800
 1741 RHAENVITPAVQWOKLEVFNAKWNFTISGTOYLAGLSTLPGNPAIASMAFTASITSP 1800
 1801 LITQNTLNLGSMVAQAOLAPPSAASAFVAGAGIAAGVSGTGLGVLDIILAGYAGVA 1860
 1801 LITQNTLNLGSMVAQAOLAPPSAASAFVAGAGIAAGVSGTGLGVLDIILAGYAGVA 1860
 1861 GALVAFKMSGEVPESTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAGVQWNRLL 1920
 1861 GALVAFKMSGEVPESTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAGVQWNRLL 1920
 1921 APASRGNHVSPTHYPESDAAARVTOILSSLIITOLLKELHOWINEDCSTPCSGWLROV 1980
 1921 APASRGNHVSPTHYPESDAAARVTOILSSLIITOLLKELHOWINEDCSTPCSGWLROV 1980
 1981 WDMICTVLTDFTKWLCKLPLPGVPEFLSCORGYKGVWRGDMQIMQTCPCQAQIAGHVK 2040
 1981 WDMICEVLSDFKWLKAKLPQLPGIFPVSCORGYKGVWRGDMQIMQTCPCQAQIAGHVK 2040
 2041 NGSMRIVGPRTCNNTWHGTFPINAYTTGCTPSPAPNYGRALWVAABEYVEVTRVGDHF 2100
 2041 NGTMRIVGPTRCNRMSGTFPINAYTTGCTPSPAPNYGRALWVAABEYVEVTRVGDHF 2100
 2101 YVTGMTDNLKCPQVAPABEFTTEVDGVRIHRVAPACKELLREDVTFOVLNOYLVGSQL 2160
 2101 YVSGMTDNLKCPQVAPABEFTTEVDGVRIHRVAPACKELLREDVTFOVLNOYLVGSQL 2160
 2161 PCBPEDVTVLTSMITDPSHITAEAKRLARGSPSLASSASQSLASAPSKATCTTHD 2220
 2161 PCBPEDVTVLTSMITDPSHITAEAKRLARGSPSLASSASQSLASAPSKATCTTHD 2220
 2221 SPDALEIANLLWRQEMGNITRVESENKVVILDSFPLHARGDEREISVAABILKRSKR 2280
 2221 SPDALEIANLLWRQEMGNITRVESENKVVILDSFPLHARGDEREISVAABILKRSKR 2280
 2281 FPGALPWARPDYNNPILLESKDPDVPVPHVGCPLPPTKAPPIPPRRKRTVVLTESNV 2340
 2281 FAPALPVWARPDYNNPILLESKDPDVPVPHVGCPLPPTKAPPIPPRRKRTVVLTESNV 2340
 2341 SSALAEIATKTFGSGSSAVDSGTATVLPOLASDDGKSGDVEVSYSMPLEGECDPDL 2400
 2341 STALAEIATKTFGSGSSAVDSGTATVLPOLASDDGKSGDVEVSYSMPLEGECDPDL 2400
 2401 SDGWSSTVSSGADTEDVCCSMYSWTGALITPCAAEESKLPINPLSNSLLRHHNLYST 2459
 2401 SDGWSSTVSSGADTEDVCCSMYSWTGALITPCAAEESKLPINPLSNSLLRHHNLYST 2459
 2460 TSRSASLROKKTFTDRLOVLDHVRDYLKEMKAKASTVKAKLLSIEACKLTPHSAKSK 2519
 2461 TSRSASLROKKTFTDRLOVLDHVRDYLKEMKAKASTVKAKLLSIEACKLTPHSAKSK 2519
 2520 FGKAKDVRMLSSRAVNIHRSWEDLLEDTEPTIDTTIMAKSEVFCVQPEKGGKPARLI 2579
 2520 FGKAKDVRMLSSRAVNIHRSWEDLLEDTEPTIDTTIMAKSEVFCVQPEKGGKPARLI 2579
 2580 VFPDLGVVCEKMALYDWWSTLPOAVMGSSYGFQYSPKQRFVFLVNTWKSCKCMGFSYD 2639
 2580 VFPDLGVVCEKMALYDWWSTLPOAVMGSSYGFQYSPKQRFVFLVNTWKSCKCMGFSYD 2639
 2640 TRCFDSTVTSBIDREBESIOCCDLAEPARQAIRSLTERLYIGGPLTNSKXGNCYRRCR 2699
 2640 TRCFDSTVTSBIDREBESIOCCDLAEPARQAIRSLTERLYIGGPLTNSKXGNCYRRCR 2699
 2700 TRCFDSTVTSBIDREBESIOCCDLAEPARQAIRSLTERLYIGGPLTNSKXGNCYRRCR 2700
 2700 TRCFDSTVTSBIDREBESIOCCDLAEPARQAIRSLTERLYIGGPLTNSKXGNCYRRCR 2700
 2701 ASGVLITSCGNLTICYLKATAACRAAKIQCTMLVNGDDLVVICESAGTOEDAAALRAFT 2759
 2701 ASGVLITSCGNLTICYLKATAACRAAKIQCTMLVNGDDLVVICESAGTOEDAAALRAFT 2759

QY 2760 EAWTRYSAPPQPOPEYDELEITSCSSNVSAHDASGRVYVLTSDPTPLARAWEYA 2819
 DB |||||
 QY 2761 EAWTRYSAPPQPOPEYDELEITSCSSNVSAHDAGGRVYLTSDPTPLARAWEYA 2820
 DB |||||
 QY 2820 RHTPINSWLGNIIMWAPTILWARMILMTHFFSIIAQOLEKALDCCQIYGACYSIEPLDLP 2879
 DB |||||
 QY 2821 RHTPVNSWLGNIIMWAPTILWARMILMTHFFSVLIARDQLEALNCBIYGACYSIEPLDLP 2880
 DB |||||
 QY 2880 QIIERLHGSAFTLHYSPOEINRVASCLKLGVPPIRWRHRARSVRAKLISQCGRAAT 2939
 DB |||||
 QY 2881 PIQRLHGLSAFSLHYSPOEINRVAAACLKLGVPPLRAWRHRARSVRARLLSRGAAAI 2940
 DB |||||
 QY 2940 CGRYLENWARTKLTPTPAASQDLGWNFVAGYSGGDIYHLSRARPRWPELCLLILS 2999
 DB |||||
 QY 2941 CGXYLENWARTKLTPTPAAGRLDLSGNFTAGYSGGDIYHSVSHPRNWFCLLLLA 3000
 DB |||||
 QY 3000 VGVGIYLLPNR 3010
 DB |||||
 QY 3001 AGVGIYLLPNR 3011
 DB |||||
 RESULT 11
 US-09-747-419-20
 ; Sequence 20, Application US/09747419
 ; Patent No. US2002015582A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lemon, Stanley
 ; APPLICANT: Yi, Minkyung
 ; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
 ; FILE REFERENCE: 265.0007 0101
 ; CURRENT APPLICATION NUMBER: US/09/747,419
 ; CURRENT FILING DATE: 2000-12-23
 ; PRIOR APPLICATION NUMBER: US 60/171,909
 ; PRIOR FILING DATE: 1999-12-23
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 20
 ; LENGTH: 3011
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Polyprotein
 US-09-747-419-20
 Query Match 88.2%; Score 14124.5; DB 9; Length 3011;
 -Best Local Similarity 85.5%; Pred. No. 0;
 Matches 2573; Conservative 218; Mismatches 219; Indels 1; Gaps 1;
 QY 1 MSTNPKPQRTKRTNRRPQDVKPPGGQIVGGYVLLPRRGPRLGVRAATKASERSQPRG 60
 DB |||||
 QY 61 RROPIKARPEGRAPQGPVPLVYNEG-GWAGWLLSPRGSPSGPPTDPRRSRNLG 120
 DB |||||
 QY 61 RROPIKARPEGTWQGPVPLVYNEGCGWAGWLLSPRGSPSGPPTDPRRSRNLG 120
 DB |||||
 QY 121 KVIDTLTCGFADLMGYIPLVCAPLGAARALAHGVVRVLEDDGVNATGNLPGCSFSIFLLA 180
 DB |||||
 QY 181 LLSCLTTPASAYEVNRVSGIYHVYNDGNSNSIIVVEAADVIMHTPGCVPCVOEGNSRCWV 240
 DB |||||
 QY 181 LLSCLTTPASAYQVRNSGLYHVYNDGNSNSIIVVEAADAILHTPGCVPCVREGNASRCWV 240
 DB |||||
 QY 241 ALTPTLAARNASVPTTIRRHVDLLAGTAAPCSAWYGLDLCGSIPLYSQLTFFSPRRHT 300
 DB |||||
 QY 241 AVTPTVATROKLPPTQURRHIDLLVGSATLCSALYVGDLCGSVFLVQQLTFFSPRRHT 300
 DB |||||
 QY 301 VQDCNCSITPGHVSCHRMAMWMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGLAGLA 360
 DB |||||
 QY 301 TQDCNCSITPGHITGHMAMWMMNWSPTAALVVAQLLRILPQAIMDMTAGAHWGLAGIA 360
 DB |||||
 QY 361 YYSVGVNWKVLIIVALLPAGVDGETHTTGRVAGHTTSGFTSLFSSGASQKIQLVNTNGSW 420
 DB |||||

DB |||||
 QY 361 YYSVGVNWKVLIIVALLPAGVDAETHVTGGNAGRTTAGLVGLLTPGAKQNTQLINTNGSW 420
 DB |||||
 QY 421 HINTALNCNDSLQTFPAALFYAHKFNSSGCPERMASCRPIDWFAOQGWGPTTTPKNSS 480
 DB |||||
 QY 421 HINTALNCNCSLNTGWLAGFYQHKFNSSGCPERLASCRRLTDFPAQGWGPTISVANGSGL 480
 DB |||||
 QY 481 DORPYCFMYFARPCGVWPAASQVCGPVYCTPTSPVVVGTTRDSGVFTYSWGENETDVMNLN 540
 DB |||||
 QY 481 DERPYCFMYFARPCGIVPAKSVCGPVYCTPTSPVVVGTTRDSGAFTYSWGANDTDVFLN 540
 DB |||||
 QY 541 NTRPQGNWFGCTWNNSTGFTKTCGGPPCNTGGVGNRPLICTDCFRKHPKPAATYKCSG 600
 DB |||||
 QY 541 NTRPPLGNWFGCTWNNSTGFTKVCAGPPCVTGGVGNNTLLCPTCFRKHPEATYSRCSG 600
 DB |||||
 QY 601 PKLTPRCLVDYPRYLWHYPCLTNFSIFKVRMTVGVGEHRLNAACNWTGERCNLEDRDS 660
 DB |||||
 QY 601 PWITPRCWDYPRYLWHYPCLTNYTIFKVRMTVGVGEHRLNAACNWTGERCNLEDRDS 660
 DB |||||
 QY 661 ELSPLLLSTTRWQILPCAFITLTPALSTGLIHQNIQVQVLYGVGSAPFAIKWEYIL 720
 DB |||||
 QY 661 ELSPLLLSTTQWQLPCSFITLTPALSTGLIHQNIQVQVLYGVGSITASWAIKWEYV 720
 DB |||||
 QY 721 LIPILLADARVCACIMMILLTAQAPAALENLVINAASVAGAHGILSLPVPFCAMWYIKG 780
 DB |||||
 QY 721 LIPILLADARVCSCLMMMLLSQAEAALENLVINAASLAGTHGLSVFLVPFCFAMWYIKG 780
 DB |||||
 QY 781 RLAPGAAYAFYGVWPLLLILLALPPRAYALDREMAASCGGAVLVGLVFTLSPYKVFIL 840
 DB |||||
 QY 781 RWVPGAVVALYGNWPLLLILLALPORAVALDEVAASCGGVLVGLMALTLSPPYKYRIS 840
 DB |||||
 QY 841 RLIMWLQYFITRAEAHMQVWVPLNVGRGRDAIILLTCAVHPLELFDITKLLAILGLPIL 900
 DB |||||
 QY 841 WCMWMLQYELTRVEAQHLVWVPLNVGRGRDAVILLMVCVHPLELFDITKLLAILFGPLW 900
 DB |||||
 QY 901 VLOAGTTRVYFVRAQGLIRACHLVKAGCHYQVFMKLGALGCTVYVNHLPURDWA 960
 DB |||||
 QY 901 ILQASLLKVPYFVRVQGLIRICALARKIAGGHVQVMAIILKALGTGYVYVNHLPURDWA 960
 DB |||||
 QY 961 HAGLRDLAVAVEPVVPSAMETKVITWADTAACGDIILGLPVARSARKEIFLGPADSLG 1020
 DB |||||
 QY 961 HNGLRDLAVAVEPVVPSRMETKLIITWADTAACGDIILGLPVARSARKQEIILGPADGWS 1020
 DB |||||
 QY 1021 QGWRLIAPITAYSQQTGRVGLGCIITSLTGRDNQVGEVQVWVSTATQSLAFCINGVCWT 1080
 DB |||||
 QY 1021 KGWRLIAPITAYAAQQTGRVGLGCIITSLTGRDNQVGEVQVISTATQTLAFCINGVCWT 1080
 DB |||||
 QY 1081 VYHGAGSKTLAGKPGPIQNTVNDVLDLVGKQAPPGARSMTPCSGSSDLVLTTRHADVI 1140
 DB |||||
 QY 1081 VYHGAGTRITASPGGPVIOQNTVNDQDLVGNWPAQPSRSLITPCTCGSSDLVLTTRHADVI 1140
 DB |||||
 QY 1141 PVRRRGDSRGSLLSPRVSYLEKSGSGPLLCPSGHVVGVRFAAVCTRGVAKAVDFIPVES 1200
 DB |||||
 QY 1141 PVRRRGDSRGSLLSPRISYLEKSGSGPLLCPSGHAVALFRAAVCTRGVAKAVDFIPVEN 1200
 DB |||||
 QY 1201 MTTMTSPVFTDNSTPAPVQFQVLAHLHAPTCGSGSTKVPAAYAAQGVKVLVNPESVA 1260
 DB |||||
 QY 1201 LGTMTSPVFTDNSTPAPVQSFQVLAHLHAPTCGSGSTKVPAAYAAQGVKVLVNPESVA 1260
 DB |||||
 QY 1261 TLGFGAYMSKAGHIDNIRITGVRTITGGSIITYSTYKFLADGGCGSGGAYDIIICDECHS 1320
 DB |||||
 QY 1261 TLGFGAYMSKAGHVDNIRITGVRTITGGSPITYSTYKFLADGGCGSGGAYDIIICDECHS 1320
 DB |||||
 QY 1321 TDSITLIGITGVLCQAEATAGARLVILATAPPGSVTVTPHPNIEEIGLSNNGEIPFYGKAI 1380
 DB |||||
 QY 1321 TDATSILIGITVLDQAEATAGARLVILATAPPGSVTVTPHPNIEEVALSTTGTPTFYGKAI 1380
 DB |||||
 QY 1381 PIFAIKGRHLIFCHSKKKCDELAALKTGLINAVAYRGLDYSVTPPTGDDVVVATDAL 1440
 DB |||||
 QY 1381 PLEVIGKGRHLIFCHSKKKCDELAALKTGLINAVAYRGLDYSVTPGDDVVVATDAL 1440
 DB |||||
 QY 1441 MTGFTGDFSDVDPNCTVCTQVDFPSLDPTFTTBTTPQDAVRSQRRGTGRGRGIYR 1500
 DB |||||

1441 MTGFTGDFDSVIDCNCVTOTVDFSLDPTFTTETTLPODAVSRTOQRGRGKPGCIYR 1500
 1501 FVTPGPRPSGNFDSVLCCEYDAGCANYELTPASTSVRLRAYLNTPLGLPCODHLEWES 1560
 1501 FVAPGPRPSGNFDSVLCCEYDAGCANYELTPASTSVRLRAYLNTPLGLPCODHLEWEG 1560
 1561 VFTGLTHIDAHFLSOTKQAGNFPYLVAYQATVCARAQAPPPSDQWKKCLIRLKPITHG 1620
 1561 VFTGLTHIDAHFLSOTKQAGNFPYLVAYQATVCARAQAPPPSDQWKKCLIRLKPITHG 1620
 1621 PTPLLYRLGAVONEVILTHPTTKYIMACMSADLEVTSTWVNGVLAALAAAYCLTTGSV 1680
 1621 PTPLLYRLGAVONEVILTHPTTKYIMACMSADLEVTSTWVNGVLAALAAAYCLTTGSV 1680
 1681 VIVGRIILSGPAPVVPREVLYQRFDEMEECASOLPYIEQGMOLAEQPKQKALGLTOTAT 1740
 1681 VIVGRIILSGPAPVVPREVLYQRFDEMEECASOLPYIEQGMOLAEQPKQKALGLTOTAT 1740
 1741 KQABAAAAPVVEKSWRALETFWAKHWNFIISGIVYLAGLSTLPGNPAIASLMAFTASITSP 1800
 1741 KQABAAAAPVVEKSWRALETFWAKHWNFIISGIVYLAGLSTLPGNPAIASLMAFTASITSP 1800
 1801 LTTQNTLLFNILGWAQAALAPPASAFVAGIAGAAVSGISGLKVLVDILAGYGAGVA 1860
 1801 LTTQNTLLFNILGWAQAALAPPASAFVAGIAGAAVSGISGLKVLVDILAGYGAGVA 1860
 1861 GALVAFKVMGSEVPSTEDVNLPAILSPGALVGVWCAAILRRHVGPGEAGVQWMMNELI 1920
 1861 GALVAFKVMGSEVPSTEDVNLPAILSPGALVGVWCAAILRRHVGPGEAGVQWMMNELI 1920
 1921 AFASRGNHVSPTHYVPESDAARVTQIILSSLTITQLKRLHWNEDCSTPCSGSWLRDV 1980
 1921 AFASRGNHVSPTHYVPESDAARVTQIILSSLTITQLKRLHWNEDCSTPCSGSWLRDV 1980
 1981 WDNICTVITDFKWLQSKLPLPGVPLSLCORGKGVARGDGTIMQTTCPCAQIAGHVK 2040
 1981 WDNICTVITDFKWLQSKLPLPGVPLSLCORGKGVARGDGTIMQTTCPCAQIAGHVK 2040
 2041 NGSMEIVGPTCSNTHGTFPPNAYTTGCTPSAPNYSRALRVAAEYVBTVRVGFH 2100
 2041 NGSMEIVGPTCSNTHGTFPPNAYTTGCTPSAPNYSRALRVAAEYVBTVRVGFH 2100
 2101 YVTGMTTQNVKCPQVPAPEFFTEVDGVRHLRYAPACKPLREDVTQVGLNQYLVGSOL 2160
 2101 YVTGMTTQNVKCPQVPAPEFFTEVDGVRHLRYAPACKPLREDVTQVGLNQYLVGSOL 2160
 2161 PCBPDPVTVLISMLTDPHITAEAKRLAGSPSLASSASOLASPLKATCTTHHD 2220
 2161 PCBPDPVTVLISMLTDPHITAEAKRLAGSPSLASSASOLASPLKATCTTHHD 2220
 2221 SPDAELIEANLLWRQEMGNITRVESENKVILDSPEPLHARGDEREISVAAEIILFKSRK 2280
 2221 SPDAELIEANLLWRQEMGNITRVESENKVILDSPEPLHARGDEREISVAAEIILFKSRK 2280
 2281 FFSALPIWAPDYNPPLLESKDDPDVVPVHGCPPLPTKAPPIPPPRKRTVWLTESV 2340
 2281 FFSALPIWAPDYNPPLLESKDDPDVVPVHGCPPLPTKAPPIPPPRKRTVWLTESV 2340
 2341 SSALAPLAKTTCSSGSSAVDSTATAPDLASDDGKSDVESYSMPLEGECDPDL 2400
 2341 SSALAPLAKTTCSSGSSAVDSTATAPDLASDDGKSDVESYSMPLEGECDPDL 2400
 2401 SDGSWTSVGEA-SEDVVCCMSYWTGALITPCAAEEGKLPINP-NSNLLRHHNMYAT 2459
 2401 SDGSWTSVGEA-SEDVVCCMSYWTGALITPCAAEEGKLPINP-NSNLLRHHNMYAT 2459
 2460 TSRSASLRQKKVTFDRLOVLDHYRDLKEMKAKATVAKLLSIEACKLTPPHSAKSK 2519
 2460 TSRSASLRQKKVTFDRLOVLDHYRDLKEMKAKATVAKLLSIEACKLTPPHSAKSK 2519
 2519 FGAKDVRNLSRANVHRSWEDLLETTETIDITIMAKSVFCVQPEKGGKPARLI 2579
 2519 FGAKDVRNLSRANVHRSWEDLLETTETIDITIMAKSVFCVQPEKGGKPARLI 2579
 2580 FGAKDVRNLSRANVHRSWEDLLETTETIDITIMAKSVFCVQPEKGGKPARLI 2580

2580 VFPLGVRVCEKMAVYDVSTLPLQAVNGSSYGPQYSPKORVEFLVNTWKSKKPMGFSYD 2639
 2581 VFPLGVRVCEKMAVYDVSTLPLQAVNGSSYGPQYSPKORVEFLVNTWKSKKPMGFSYD 2640
 2640 TRCFDSTVTSRDETRVESIYOCCLAPPEARQATRSALTERLYIGGPLTNSKGQCGYRRCR 2699
 2641 TRCFDSTVTSRDETRVESIYOCCLAPPEARQATRSALTERLYIGGPLTNSKGQCGYRRCR 2700
 2700 ASGVLTTSCTNTLTCTVLKATAACRAAKLQDCTMLVNGDDLVVICSAGTQEDAAALRAFT 2759
 2701 ASGVLTTSCTNTLTCTVLKATAACRAAKLQDCTMLVNGDDLVVICSAGTQEDAAALRAFT 2760
 2760 EAMTRYAPPDGPDPPEYDLELTITSSNSVSVHADSKGVYVLTDPDPTPLARAAWETA 2819
 2761 EAMTRYAPPDGPDPPEYDLELTITSSNSVSVHADSKGVYVLTDPDPTPLARAAWETA 2820
 2820 RHTPINSWLGNIIMVAPTLWARMILMTHFFSIIILAQOLEKALDCQIYGACYSIEPDLIP 2879
 2821 RHTPINSWLGNIIMVAPTLWARMILMTHFFSIIILAQOLEKALDCQIYGACYSIEPDLIP 2880
 2880 QIETRLHGLSAFTLHSYSPGEINRVASCLRLKLGVPPLTWRHRARSVRKLLSGGGAAT 2939
 2881 QIETRLHGLSAFTLHSYSPGEINRVASCLRLKLGVPPLTWRHRARSVRKLLSGGGAAT 2940
 2940 CGRYLFNWAVRTKLTPIPAASQDLSCWFWAGYSGDDIYHLSRABDPWFPLCLLLS 2999
 2941 CGRYLFNWAVRTKLTPIPAASQDLSCWFWAGYSGDDIYHLSRABDPWFPLCLLLS 3000
 3000 VGVGYYLLPNR 3010
 3001 VGVGYYLLPNR 3011

RESULT 12
 ; Sequence 14, Application US/10189359
 ; Publication No. US20040039187A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SANGAR, DAVID V.
 ; APPLICANT: LEMON, STANLEY M.
 ; TITLE OF INVENTION: Chimeric GB Virus B (GBV-B)
 ; FILE REFERENCE: US/10/189,359
 ; CURRENT APPLICATION NUMBER: US/10/189,359
 ; CURRENT FILING DATE: 2002-07-03
 ; PRIOR APPLICATION NUMBER: 10/189,359
 ; PRIOR FILING DATE: 2002-07-03
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 14
 ; LENGTH: 3011
 ; TYPE: PRT
 ; ORGANISM: Hepatitis C virus
 ; US-10-189-359-14

Query Match 88.2%; Score 14124.5; DB 12; Length 3011;
 Best Local Similarity 85.5%; Pred. No. 0;
 Matches 2573; Conservative 218; Mismatches 219; Indels 1; Gaps 1;

QY 1 MSTNPKPQKTKRNTNRPPQDVKPPGGQIVGVVLLPRRGPRGLGVATRKASERSQPRG 60
 DB 1 MSTNPKPQKTKRNTNRPPQDVKPPGGQIVGVVLLPRRGPRGLGVATRKTSERSQPRG 60
 QY 61 RRQPIKARRPGRAWAOPGVWPPLYGNEGLGWAGWLLSPRGSRSWQPTDPRRSRNLG 120
 DB 61 RRQPIKARRPGRITWAQEGYWPPLYGNEGCWAGWLLSPRGSRSWQPTDPRRSRNLG 120
 QY 121 KVIDLTTCFADLMGYIPLVGAAPLGAARALAHGVTVLEDGVNYATGNLPGCSSEIFLLA 180
 DB 121 KVIDLTTCFADLMGYIPLVGAAPLGAARALAHGVTVLEDGVNYATGNLPGCSSEIFLLA 180
 QY 181 LLSCLTIPASAEVENVNSGIYHTVNDSCNSSIYIEAADVIMHTPTGCVPCVQGNSSRCWV 240

Db 181 LUSCLVPA5AYOVRNSGHIWTDNCPASSIYVEAADAILHTPCVPCVRGNA5RCWV 240
QY 241 ALPTLAARNA5VFTTIRRHVDLLYGTAAFC5AMVYGDLCG5IFLV50:FTPSRRHET 300
Db 241 AVPTVATRDGKLPFTQTLRSHIDLLYGSATLCSALYVGLCG5VFLVGQFLTSPRRHWT 300
QY 301 VQDCNCSIYGHVSHHMANDMNNSPTTALVUSQILRIPOAVDMVAGAHGVLGAGLA 360
Db 301 TQDCNCSIYGHGHTGHRMANDMNNSPTAALVVAQLLRIPQAIMDIAGAHGVLGAGLA 360
QY 361 Y5VMVGNKAKVILVAILFACVDGETHTGRVAGHTTSGFTSLFSSGASQKIQLVNTINGSW 420
Db 361 YFSVMGNKAKVILVLLFAGVDAETHVTGNGAGETTAGLVGLLTPGKAKQNIQLININGSW 420
QY 421 HINRTALNCNDSLOTGFFAALFYAHKNSGSCPERMASCRPIUWFAOGWGPITYTENS5 480
Db 421 HINSTALNCNENLNTGMLAGFYCHKNSGSCPERLASCRRLTDFAQGWGPI5YANGSGL 480
QY 481 DORPYCHYAPRPGVTPASQVCGPVYCFTPSPVAVGTTDRSGVPTYSWGENETDUMLLN 540
Db 481 DERPYCHYPRPGIYFPAKSVCQPVYCFTPSPVAVGTTDRSGAPTYSWGANDIDVFLN 540
QY 541 NTRPPQGNWFCTWNSGTFTKTCGGPPONIGGVGNRTLICPTDCPRKHPEATYTKCGSG 600
Db 541 NTRPPLGNWFCTWNSGTFTKVCGAPPVIGGVGNNTLLCPTDCPRKHPEATYSROGSG 600
QY 601 PWLTPRCLVDYFYRLMHYPCNLNFSIKVRMYGVGVZHLNAAQNWTRGERCNLEDDRS 660
Db 601 PWITPRCMVDYFYRLMHYPCNTYTIKVRMYGVGVZHLNAAQNWTRGERCNLEDDRS 660
QY 661 ELSPLLLSTTBWOLILPCAFITLPAISLGLHLHONIVDVQVLYGVGSAFVSFALKWBYIL 720
Db 661 ELSPLLLSTTBWOLILPCAFITLPAISLGLHLHONIVDVQVLYGVGSAFVSFALKWBYIL 720
QY 721 LLLFLLADARVCACIMWMLLIQAEEAALLENVLNAA5VAGAHGIL5FLVFCAAWYIKG 780
Db 721 LLLFLLADARVCSCIMWMLLIQAEEAALLENVLNAA5LAGTHGLV5PLVFFCFPAWYIKG 780
QY 781 RLAPQAAVAFYGVWPLLLLLALPRAYALDREMAASCGGVALGVL5LTPSPYKYVFLT 840
Db 781 RWVPQAVYALYGMWFLLLLLALPORAYALDEVAASCGGVVL5GLMALTI5PYKYRYS 840
QY 841 RLIMWLOYEITRAEAMQVWVPEPLNVGRGDALILLTCAVHEPILFDITKLLALLGLPLM 900
Db 841 WCMWMLQFLTRVEAQLHWVPLNVGRGDALILLMVCVHPTLVDITKLLALLGLPLM 900
QY 901 VLQAGITRVPYFVRAQGLIRACMLYKXVAGGHYQVWFNKI5GALTGTVYVNHLPRLDWA 960
Db 901 ILQASLLXVPYFVRYOGLLRIICALARKIAGGHYQVMAIILK5GALTGTVYVNHLPRLDWA 960
QY 961 HAGLRLDIAVAVEPVVFSAMEKVIITWGDATACGDIILGLPUSARGKEIFLGPAD5LEG 1020
Db 961 HNGLEDIAVAVEPVVFSAMEKVIITWGDATACGDIINGLPVSARGKEILGPADGVS 1020
QY 1021 OGWRLLAPITAYSQOTRGVLGCIITSLTGRDNQVGEVQVYSTAQ5FLATPCINGVCWT 1080
Db 1021 KGRWLLAPITAYAOQTRGLGCIITSLTGRDNQVGEVQVYSTAQ5FLATPCINGVCWT 1080
QY 1081 VYHAGSKTLAGEKGPITQMYTNVDLIDVWQAPPGARSNTFC5GSSDLYLVTRHADVI 1140
Db 1081 VYHAGATRIASGPKGPVIQMYTNVDQDLVWGPAPQGSRLT5CTCGSSDLYLVTRHADVI 1140
QY 1141 PYRRRGDSGSL5REPPV5YLGSSG5P5LLCP5GHVGVFRAAVCTRGVAKAVDFIPVES 1200
Db 1141 PYRRRGDSGSL5R5P5IS5TLGSSG5P5LLCP5GHVGVFRAAVCTRGVAKAVDFIPVEN 1200
QY 1201 METTMR5PVTDNSTPPAVPQTFQVAHL5EAP5GSK5TKVPAAYAAQ5KVLVLN5P5VAA 1260
Db 1201 LGTTMR5PVTDNSTPPAVPQ5QVAHL5EAP5GSK5TKVPAAYAAQ5KVLVLN5P5VAA 1260
QY 1261 TLF5G5Y5K5AG5IDN5IR5G5V5IT5TG5IT5T5Y5G5K5L5AG5G5G5Y5D5I5I5C5E5H5 1320

Db 1261 TLF5G5Y5K5AG5IDN5IR5G5V5IT5TG5IT5T5Y5G5K5L5AG5G5G5Y5D5I5I5C5E5H5 1320
QY 1321 TDSTIILG5ITVLDQ5ETAG5ARLV5L5AT5PP5SVTV5PH5NI5E5I5GL5N5GE5IP5FY5K5AI 1380
Db 1321 TDSTIILG5ITVLDQ5ETAG5ARLV5L5AT5PP5SVTV5PH5NI5E5I5GL5N5GE5IP5FY5K5AI 1380
QY 1381 PIEAIK5G5R5HL5IFCH5KK5C5DE5LA5AK5T5GL5IN5AV5Y5R5GL5DV5VIP5IG5DV5V5AT5DAL 1440
Db 1381 PLEVIK5G5R5HL5IFCH5KK5C5DE5LA5AK5T5GL5IN5AV5Y5R5GL5DV5VIP5IG5DV5V5AT5DAL 1440
QY 1441 MTG5FG5D5P5SV5TD5C5NT5VT5Q5T5V5D5F5L5D5P5FT5T5T5V5D5V5R5O5R5G5RT5G5R5GI5YR 1500
Db 1441 MTG5FG5D5P5SV5TD5C5NT5VT5Q5T5V5D5F5L5D5P5FT5T5T5V5D5V5R5O5R5G5RT5G5R5GI5YR 1500
QY 1501 FVTP5R5P5G5M5D5SS5VL5C5Y5D5AG5CAW5YEL5T5P5A5T5SV5KL5AY5L5NT5P5G5P5V5O5D5H5LE5F5W5E5 1560
Db 1501 FVAP5R5P5G5M5D5SS5VL5C5Y5D5AG5CAW5YEL5T5P5A5T5SV5KL5AY5L5NT5P5G5P5V5O5D5H5LE5F5W5E5 1560
QY 1561 VFT5GL5TH5IDA5H5FL5S5Q5K5AG5DN5FP5YL5V5AY5Q5V5CAR5AQ5APP5PS5W5Q5M5K5CL5RL5K5PT5L5HG 1620
Db 1561 VFT5GL5TH5IDA5H5FL5S5Q5K5AG5DN5FP5YL5V5AY5Q5V5CAR5AQ5APP5PS5W5Q5M5K5CL5RL5K5PT5L5HG 1620
QY 1621 PT5PL5LY5EL5G5AV5Q5NE5V5L5TH5PI5TK5Y5IM5AC5S5AD5LEV5VT5TW5LV5G5V5LA5AL5AY5CL5T5GSV 1680
Db 1621 PT5PL5LY5EL5G5AV5Q5NE5V5L5TH5PI5TK5Y5IM5AC5S5AD5LEV5VT5TW5LV5G5V5LA5AL5AY5CL5T5GSV 1680
QY 1681 VI5GR5I5L5SG5K5P5AV5PD5REV5LY5Q5E5F5D5E5M5E5C5A5Q5L5PY5I5EQ5M5Q5L5A5F5Q5K5AL5G5L5L5Q5AT 1740
Db 1681 VI5GR5I5L5SG5K5P5AV5PD5REV5LY5Q5E5F5D5E5M5E5C5A5Q5L5PY5I5EQ5M5Q5L5A5F5Q5K5AL5G5L5L5Q5AT 1740
QY 1741 KQ5E5A5AP5V5B5SK5W5RA5L5E5F5W5AK5M5N5F5I5SG5I5O5Y5L5AG5L5TL5P5GN5PA5I5AS5I5MA5T5A5IT5SP 1800
Db 1741 RHA5E5IT5P5AV5Q5T5W5K5LE5V5F5W5AK5M5N5F5I5SG5I5O5Y5L5AG5L5TL5P5GN5PA5I5AS5I5MA5T5A5IT5SP 1800
QY 1801 LTT5Q5T5L5L5F5N5IL5G5W5A5A5Q5L5APP5S5A5AF5V5G5AG5I5AG5A5V5G5I5GL5K5V5L5D5I5L5AG5Y5G5A5V 1860
Db 1801 LTT5Q5T5L5L5F5N5IL5G5W5A5A5Q5L5APP5S5A5AF5V5G5AG5I5AG5A5V5G5I5GL5K5V5L5D5I5L5AG5Y5G5A5V 1860
QY 1861 GAL5V5AK5W5M5G5E5V5P5T5E5DL5V5N5L5L5P5AIL5P5G5AL5V5V5V5CA5IL5RR5H5V5G5E5G5A5V5Q5W5NR5LI 1920
Db 1861 GAL5V5AK5W5M5G5E5V5P5T5E5DL5V5N5L5L5P5AIL5P5G5AL5V5V5V5CA5IL5RR5H5V5G5E5G5A5V5Q5W5NR5LI 1920
QY 1921 AF5AS5R5G5N5H5V5P5TH5V5P5B5D5A5A5V5T5Q5L5S5L5T5I5T5O5L5K5R5L5H5O5M5I5N5E5D5C5P5C5S5G5M5L5RDV 1980
Db 1921 AF5AS5R5G5N5H5V5P5TH5V5P5B5D5A5A5V5T5Q5L5S5L5T5I5T5O5L5K5R5L5H5O5M5I5N5E5D5C5P5C5S5G5M5L5RDV 1980
QY 1981 WD5M5I5CT5V5L5T5D5FK5W5L5Q5K5L5P5L5P5GV5P5FL5C5Q5RG5Y5K5W5R5G5D5G5I5M5Q5T5C5P5C5A5Q5I5AG5HVX 2040
Db 1981 WD5M5I5CT5V5L5T5D5FK5W5L5Q5K5L5P5L5P5GV5P5FL5C5Q5RG5Y5K5W5R5G5D5G5I5M5Q5T5C5P5C5A5Q5I5AG5HVX 2040
QY 2041 NG5M5R5I5V5G5P5R5T5C5N5T5W5G5T5P5I5N5A5Y5T5G5P5C5T5P5G5P5A5N5Y5S5AL5R5V5A5E5Y5E5V5T5R5U5G5D5F5E 2100
Db 2041 NG5M5R5I5V5G5P5R5T5C5N5T5W5G5T5P5I5N5A5Y5T5G5P5C5T5P5G5P5A5N5Y5S5AL5R5V5A5E5Y5E5V5T5R5U5G5D5F5E 2100
QY 2101 YV5T5G5M5T5D5N5K5C5Q5V5AP5P5E5F5E5D5G5V5R5L5H5Y5AP5ACK5P5LL5R5E5D5V5T5Q5V5G5L5N5Q5V5L5G5S5Q5L 2160
Db 2101 YV5G5M5T5D5N5K5C5Q5I5P5P5E5F5E5D5G5V5R5L5H5Y5AP5ACK5P5LL5R5E5D5V5T5Q5V5G5L5N5Q5V5L5G5S5Q5L 2160
QY 2161 PCE5PE5P5DV5L5T5M5L5T5D5P5SH5I5A5E5R5L5ARG5P5PS5I5A5S5S5Q5L5A5P5S5L5K5AT5CT5TH5D 2220
Db 2161 PCE5PE5P5DV5L5T5M5L5T5D5P5SH5I5A5E5R5L5ARG5P5PS5I5A5S5S5Q5L5A5P5S5L5K5AT5CT5TH5D 2220
QY 2221 SP5D5AD5L5E5AN5L5L5W5Q5E5M5G5N5T5R5V5E5N5K5V5I5D5S5P5E5P5L5H5A5E5G5E5R5I5S5V5A5E5L5R5K5SRK 2280
Db 2221 SP5D5AD5L5E5AN5L5L5W5Q5E5M5G5N5T5R5V5E5N5K5V5I5D5S5P5E5P5L5H5A5E5G5E5R5I5S5V5A5E5L5R5K5SRK 2280
QY 2281 PPS5AL5P5I5W5AR5D5Y5N5P5P5L5E5S5W5K5D5P5Y5V5P5V5H5G5C5P5L5E5T5K5AP5T5K5AP5P5R5K5T5V5L5T5E5SNV 2340
Db 2281 FAR5AL5P5W5AR5D5Y5N5P5P5L5E5S5W5K5D5P5Y5V5P5V5H5G5C5P5L5E5T5K5AP5T5K5AP5P5R5K5T5V5L5T5E5SNV 2340
QY 2341 SS5AL5A5E5L5AT5K5T5F5G5S5G5S5A5V5D5S5T5A5L5P5D5L5A5S5D5G5D5K5G5D5V5E5S5M5P5P5E5G5P5D5DL 2400
Db 2341 ST5AL5A5E5L5AT5K5T5F5G5S5G5S5A5V5D5S5T5A5L5P5D5L5A5S5D5G5D5K5G5D5V5E5S5M5P5P5E5G5P5D5DL 2400

2401 SDGSWSTVSEEA-SEDVYCCSMSTYWTGALITPCAAEESKLPINPLNSLLRHHNMVYAT 2459
2401 SDGSWSTVSSGADTEDVYCCSMSTYWTGALVTPCAABEQKLPINALNSLLRHHNLVYST 2460
2460 TSRSASIROKAVFDRLOVLDHVRDLKEMKAAKASTVKAKLLSIEBACKLTPPHSAKSK 2519
2461 TSRSACORQKAVTDRLOVLDHVRDLKEMKAAKASTVKAKLLSIEBACKLTPPHSAKSK 2520
2520 FGYGAKDVRNLSSRAVNHLSVWDLLEDETPDTTIMAKSEVFCVQPKGGRKPARLI 2579
2521 FGYGAKDVRCHARKAVAHINSVWDLLEDSVTPDTTIMAKNEVFCVQPEKGRKPARLI 2580
2580 VFPDLGVRCFKMALYDVVSTLPOAVMGSSYGFQSPKORVEFLVNTWKSKKCPMGFSYD 2639
2581 VFPDLGVRCFKMALYDVVSKLPLAVMGSSYGFQSPQORVEFLVQVAKSKKTPMGFSYD 2640
2640 TRCFDSTVTESDIRVEESIYQCCDLAPARQAIRSLTERLYIGGLPLNSKQONCYRRCR 2699
2641 TRCFDSTVTESDIRVEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCYRRCR 2700
2700 ASGVLTSCGNLTICYLKATACRAAKLQDCTMLVNGDDIIVICISAGTQDAAALRAFT 2759
2701 ASGVLTSCGNLTICYLKARAACRAAGLQDCTMLVCGDDLIVICISAGVQDAAALRAFT 2760
2760 EAMTRYAPPDPPPOPEYDLELITSCSNVSVAHNDASGRVYVYLDTPPTLARAAMETA 2819
2761 EAMTRYAPPDPPPOPEYDLELITSCSNVSVAHNDAGKRVYVYLDTPPTLARAAMETA 2820
2820 RHTPNSMLGNIIMVAPTLWARMILMTHFFSILLAOEQLKALDCQIYACYSIPELDLP 2879
2821 RHTPNSMLGNIIMVAPTLWARMILMTHFFSVLIARDQLEQALNCEIYACYSIPELDLP 2880
2880 QIERLHGLSAPLTHSYSPGEINRVASCLKLVPLRTWHRARSVRAKLLSQGGRAT 2939
2881 PIQIRLHGLSAPLTHSYSPGEINRVAAACLKLVPLRAWRHARSVRARLLSRRGRAT 2940
2940 CGRYLFNVAVRKLTLPAPASQDLSCGFVAGYSGGDIYHSLRARPRWFPLCILLIS 2999
2941 CGRYLFNVAVRKLTLPAPAGRLDLSGFFAGYSGGDIYHVSSEARPRWFPLCILLIS 3000
3000 VGVGIYLLPNR 3010
3001 AGVGIYLLPNR 3011

RESULT 13
US-10-259-275-20
; Sequence 20, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
; APPLICANT: Lemon, Stanley M.
; APPLICANT: Yi, Minkyung
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
; FILE REFERENCES: 265,007 0120
; CURRENT APPLICATION NUMBER: US/10/259, 275
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US 60/171,909
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 09/747,419
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/325,236
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/338,123
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 3011
; TYPE: PRI
; ORGANISM: artificial
; FEATURES:
; OTHER INFORMATION: Polyprotein

US-10-259-275-20

Query Match 88.2%; Score 14124.5; DB 14; Length 3011;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 2573; Conservative 218; Mismatches 219; Indels 1; Gaps 1;

QY 1 MSTNPQKRTKNTNRRPQDVKFPGGQIVGGVILLPRGPRGLGVATRKASERQPRG 60
DB 1 MSTNPQKRTKNTNRRPQDVKFPGGQIVGGVILLPRGPRGLGVATRKASERQPRG 60

QY 61 RRPIKARPEGRAWAQPCYPMPLVYNEGLGWAGMLLSPRGSRPSWGPTDPRRRRNLG 120
DB 61 RRPIKARPEGRAWAQPCYPMPLVYNEGLGWAGMLLSPRGSRPSWGPTDPRRRRNLG 120

QY 121 KVLDLTCGFPADLMGYIPLVGAFLGGAARALAHGVLEDCVYATCNLPGCCSFLLA 180
DB 121 KVLDLTCGFPADLMGYIPLVGAFLGGAARALAHGVLEDCVYATCNLPGCCSFLLA 180

QY 181 LLSCLITPASAYEVNVSIGYHVNTDCNSSIYVEADVIMHTPGCVPCVQEGNSRRCW 240
DB 181 LLSCLITPASAYEVNVSIGYHVNTDCNSSIYVEADVIMHTPGCVPCVQEGNSRRCW 240

QY 241 ALTEPLAARNASVPTTIRRHVDLLVGTAAAFCSAMVYVGDLCGSIFLVSLFTSPRHET 300
DB 241 AVTPTVATRGKLPFTQRRHIDLLVGSATLCSALYVGDLCGSIFLVSLFTSPRHET 300

QY 301 VQDCNCSIYPGHVSGHRMAMMMNSPTTALVVSQLLRIPQAVVDMVAGAHWGLA 360
DB 301 TODCNCSIYPGHVSGHRMAMMMNSPTTALVVSQLLRIPQAVVDMVAGAHWGLA 360

QY 361 YYSVGNWAKVLIYALLFAGVDGTHHTTGRVAGHTTSGFTSLFSSGASQIKLVNTGWS 420
DB 361 YYSVGNWAKVLIYALLFAGVDGTHHTTGRVAGHTTSGFTSLFSSGASQIKLVNTGWS 420

QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMACSRPIDMFAQWGTITTKPNS 480
DB 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMACSRPIDMFAQWGTITTKPNS 480

QY 481 DORPYCWHYAPRPGCVGPVCPASOVCPVTCFTSPVVTGTTDRSGVPTYSGMNETDVLIN 540
DB 481 DORPYCWHYAPRPGCVGPVCPASOVCPVTCFTSPVVTGTTDRSGVPTYSGMNETDVLIN 540

QY 541 NTRPPQGNWFGCTWMNSTGFTKCGPPCNTGGVGNRTLICPTDCFRKHEATYTKCSG 600
DB 541 NTRPPQGNWFGCTWMNSTGFTKCGPPCNTGGVGNRTLICPTDCFRKHEATYTKCSG 600

QY 601 PWLTPRCVLDYFYRLWHYPCTLNFSIFKVMYVGVGVEHRLNAACNWTGERCNLEDRS 660
DB 601 PWLTPRCVLDYFYRLWHYPCTLNFSIFKVMYVGVGVEHRLNAACNWTGERCNLEDRS 660

QY 661 ELSPLLSTTEWQIILPCAFITLIPALSTGLIHLHQNIVDVQVLYGVGSAPFAIKWEVIL 720
DB 661 ELSPLLSTTEWQIILPCAFITLIPALSTGLIHLHQNIVDVQVLYGVGSAPFAIKWEVIL 720

QY 721 LFLLLADARVACACILWMLLITACAAALENILVLAASVAGAHGILSLFVFCAAWYIK 780
DB 721 LFLLLADARVACACILWMLLITACAAALENILVLAASVAGAHGILSLFVFCAAWYIK 780

QY 781 RLAPGAAYAFYGVWPIILLLLALPPRAVALDREMAASCGAVLGLVFLITSPYKVFIL 840
DB 781 RLAPGAAYAFYGVWPIILLLLALPPRAVALDREMAASCGAVLGLVFLITSPYKVFIL 840

QY 841 RLIMLQYPTITRAEAMQWVPLNVRGGRDAIILITCAVHPELIFDITKLLALITGLIM 900
DB 841 RLIMLQYPTITRAEAMQWVPLNVRGGRDAIILITCAVHPELIFDITKLLALITGLIM 900

QY 901 VLOAGITRVPYFVRAOGLTRACMLKVKVAGGVQVFMKLGALTCTYVYNHLTPLRWA 960
DB 901 VLOAGITRVPYFVRAOGLTRACMLKVKVAGGVQVFMKLGALTCTYVYNHLTPLRWA 960

QY 961 HAGLRDLAVAVEPVPFSAMETKITWTGADTAACGDIILGLPVSARKGKIFLGPADLSG 1020
DB 961 HAGLRDLAVAVEPVPFSAMETKITWTGADTAACGDIILGLPVSARKGKIFLGPADLSG 1020

1021 QGWRLLADITAYSQOTRGVLGCIITSLGRDNQNOVEGEVOVUSTATQSFATCINGVCWT 1080
1021 KGRWLLADITAYAQOTRGLGCIITSLGRDNQNOVEGEVOVUSTATQSFATCINGVCWT 1080
1081 VYHGAGSXTLAGPKGPIQMTYTNVDLVLGWAQAPGARSMTPCSCGSDLYLVTRHADVI 1140
1081 VYHGAGTRIASPKGPVQMTYTNVDLVLGWAQAPGARSMTPCSCGSDLYLVTRHADVI 1140
1141 PVRRGDRGSLSPRPVSYLKGSGGGLLCPGSHVGVFRAAVCTRGVAKAVDPIVRS 1200
1141 PVRRGDRGSLSPRPVSYLKGSGGGLLCPGSHVGVFRAAVCTRGVAKAVDPIVRS 1200
1201 METTWSPVFTDNTSPVAPVQFQVAHLHAPTGSKSTKVPAAVAAQGYKVLVLPNSVAA 1260
1201 LGTTWSPVFTDNTSPVAPVQFQVAHLHAPTGSKSTKVPAAVAAQGYKVLVLPNSVAA 1260
1261 TLGFGAYMSKAGHIDPNIITGGSIIYTSYTGKFLADGGCGGAYDIIICDECHS 1320
1261 TLGFGAYMSKAGHIDPNIITGGSIIYTSYTGKFLADGGCGGAYDIIICDECHS 1320
1321 TDSITLIGTGLVLDQAFAGABLVLATATPGSVTVPHNIEEIGLSNNGEIPFGKAI 1380
1321 TDSITLIGTGLVLDQAFAGABLVLATATPGSVTVPHNIEEIGLSNNGEIPFGKAI 1380
1381 PTEAIKGGHLLFCFSKKKDELAALGLGLNNAVAYRGLDVSVPPIGDVVVAATDAL 1440
1381 PTEAIKGGHLLFCFSKKKDELAALGLGLNNAVAYRGLDVSVPPIGDVVVAATDAL 1440
1441 MTGFTGDFDSVDCNTCVTQTVDFSLDPTFTTETTTVPQDAVSRSORRGTRGRSGIYR 1500
1441 MTGFTGDFDSVDCNTCVTQTVDFSLDPTFTTETTTVPQDAVSRSORRGTRGRSGIYR 1500
1501 FVTGPRBGMEDSSVLCSCYDAGCAYELTPARTSVRLAYLNTPLGVCQDHLBZMES 1560
1501 FVTGPRBGMEDSSVLCSCYDAGCAYELTPARTSVRLAYLNTPLGVCQDHLBZMES 1560
1561 VFTGLTHIDAHFLSOTKQAGDNFYLAVQATVCARAQAPPPSDQWKKLIRLKTFLHG 1620
1561 VFTGLTHIDAHFLSOTKQAGDNFYLAVQATVCARAQAPPPSDQWKKLIRLKTFLHG 1620
1621 PTELLYRLGAVONEVILTHITKYLMACMSADLEVTVSTWVLVGGVLAALAYCLTGTGV 1680
1621 PTELLYRLGAVONEVILTHITKYLMACMSADLEVTVSTWVLVGGVLAALAYCLTGTGV 1680
1681 VIYGRILISKPAVDPREVLYOSSDEMECASOLPYIBOGMOLASQFKOKALGLLQTAT 1740
1681 VIYGRILISKPAVDPREVLYOSSDEMECASOLPYIBOGMOLASQFKOKALGLLQTAT 1740
1741 KQAEAAAPVSVESKRALETFAWKHMNFISGIOYLAGLSLPGNPAIASLMAFTASITSP 1800
1741 KQAEAAAPVSVESKRALETFAWKHMNFISGIOYLAGLSLPGNPAIASLMAFTASITSP 1800
1801 LTTQNTLLFNILGWVAAQALAPSAASAFVAGIAGNAGVCSIGLKVLDILAGYGAV 1860
1801 LTTQNTLLFNILGWVAAQALAPSAASAFVAGIAGNAGVCSIGLKVLDILAGYGAV 1860
1861 GALVAPKVMGVEPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGFPGGAVQWMMRLI 1920
1861 GALVAPKVMGVEPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGFPGGAVQWMMRLI 1920
1921 AFASRGNHVSFTHYVPESDAAARVTOILSSITITQLLKRHLQWINEDECSTPCSGSWLRDV 1980
1921 AFASRGNHVSFTHYVPESDAAARVTOILSSITITQLLKRHLQWINEDECSTPCSGSWLRDV 1980
1981 WDWITCTVLTDFKTLWQSKLLPRLPGVFLSCORGKGVWRGDMQTCPCGAGIAGHV 2040
1981 WDWITCTVLTDFKTLWQSKLLPRLPGVFLSCORGKGVWRGDMQTCPCGAGIAGHV 2040
2041 NGSMRIIVGPRICSNWHTGTPPINAYTTGCPSPAPNYSRALMRVAAEEYVEVTRVGDFF 2100
2041 NGSMRIIVGPRICSNWHTGTPPINAYTTGCPSPAPNYSRALMRVAAEEYVEVTRVGDFF 2100

RESULT 14

US-09-316-359-2
; Sequence 2, Application US/09916359
; Patent No. US20020034734A1
; GENERAL INFORMATION:
; APPLICANT: Veronique Barbar
; TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
; TREATING C HEPATITIS

2101 YVTGMTTNDNKKCPQCPVAPPEFFTEVDCVRUERAPACKPILLREDVTFQVGLNQVLQSQL 2160
2101 YVTGMTTNDNKKCPQCPVAPPEFFTEVDCVRUERAPACKPILLREDVTFQVGLNQVLQSQL 2160
2161 PCEPEPDVTLTSMITDPSHITATATGRLARGSPPLASSASOLSAPSKATCTTHHD 2220
2161 PCEPEPDVTLTSMITDPSHITATATGRLARGSPPLASSASOLSAPSKATCTTHHD 2220
2221 SPDAELTEANLLWEQMGGMNITRVESENKVILDSFPLHAEGDEREISVAAETILRSKR 2280
2221 SPDAELTEANLLWEQMGGMNITRVESENKVILDSFPLHAEGDEREISVAAETILRSKR 2280
2281 FPSALPIWADYNNPPLLESWKPDYVPPVHGCPLPPTKAPPIPPPRKRTVVTESNV 2340
2281 FARALPVWADYNNPPLLESWKPDYVPPVHGCPLPPTKAPPIPPPRKRTVVTESNV 2340
2341 SSALAEATKTFGSSGSAVDSGTATATPLASDDGDKSDVESYSMPPELEGPGPDL 2400
2341 SSALAEATKTFGSSGSAVDSGTATATPLASDDGDKSDVESYSMPPELEGPGPDL 2400
2401 SDGWSMTVSEEA-SDVVCSSMSYTWTCALITPCAEEESKLPIPLNSLLRHHNVAT 2459
2401 SDGWSMTVSEEA-SDVVCSSMSYTWTCALITPCAEEESKLPIPLNSLLRHHNVAT 2459
2460 TSRSASLRQKKVTPDRLOVDHTRDVLKEMKAKASTVKALLSTEEACKLTTPHSAKSK 2519
2461 TSRSACORQKKVTPDRLOVDHTRDVLKEMKAKASTVKALLSTEEACKLTTPHSAKSK 2520
2520 FGYGAKDVNLSRRVNHRSVWEDLLEDTPIDTTIMAKSEVPCVOPKGGKRPALII 2579
2521 FGYGAKDVNLSRRVNHRSVWEDLLEDTPIDTTIMAKSEVPCVOPKGGKRPALII 2580
2580 VFPDLGVRVCEKVALYDVVSKLPLAVMGSSYGFQVSPGQVVEFLVQAWKSKKTPMGFSYD 2639
2581 VFPDLGVRVCEKVALYDVVSKLPLAVMGSSYGFQVSPGQVVEFLVQAWKSKKTPMGFSYD 2640
2640 TRCFPSTVTEBDRVSESIYQCCDLAPARQAIRSLTERLYIGGLTNSKQNCQYRRCR 2699
2641 TRCFPSTVTEBDRVSESIYQCCDLAPARQAIRSLTERLYIGGLTNSKQNCQYRRCR 2700
2700 ASGVLTTSQNTLTCTYKATAACRAAKLQDCITMLVNGDDLVCESAGTQEDDAAALRAFT 2759
2701 ASGVLTTSQNTLTCTYKATAACRAAKLQDCITMLVNGDDLVCESAGTQEDDAAALRAFT 2760
2760 EAMTRYSAPPQPPQPEYDLELITSCSNVSVAHDAKGRVYLTTRDPTTPIARAWEA 2819
2761 EAMTRYSAPPQPPQPEYDLELITSCSNVSVAHDAKGRVYLTTRDPTTPIARAWEA 2820
2820 RHTPINSWLGNIIMYAPTLWABWILNTHFFSTLLAQEQLEKALDCQIYCACYSIEPDLDP 2879
2821 RHTPINSWLGNIIMYAPTLWABWILNTHFFSTLLAQEQLEKALDCQIYCACYSIEPDLDP 2880
2880 QIIRLHGLSAFTLHSHYSPEINRVASCLRLKLGVPPLTRWHRARSVRALLSQGGAAT 2939
2881 PIQRLHGLSAFTLHSHYSPEINRVASCLRLKLGVPPLTRWHRARSVRALLSQGGAAT 2940
2940 CRYLPFNWAVRKLKLTPIPAASQLDLSQWFAVSGGDIYHLSLRAPRPNWPLCLILLIS 2999
2941 CRYLPFNWAVRKLKLTPIPAASQLDLSQWFAVSGGDIYHLSLRAPRPNWPLCLILLIS 3000
3000 VGVGYILLPNR 3010
3001 AGVGYILLPNR 3011

```

; FILE REFERENCE: PMCF97-03A
; CURRENT APPLICATION NUMBER: US/09/916.359
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 09/388,874
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: 97/02,887
; PRIOR FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Virus
US-09-916-359-2

Query Match      88.2%; Score 14119.5; DB 9; Length 3011;
Best Local Similarity 85.2%; Pred. No. 0;
Matches 2564; Conservative 226; Mismatches 220; Indels 1; Gaps 1;

Qy 1 MSTNPKQKTKRNTNRRPQDKFPGGQIVGGVYLLPRRPRGLGVRAATKASRSQPRG 60
Db 1 MSTNPKQKKNKNTNRRPQDKFPGGQIVGGVYLLPRRPRGLGVRAATKTSRSQPRG 60

Qy 61 RRQIPKARRPREGRAWQPGVWBL YNNEGLGWAGWLLSPRGSPPSGPTDPRRSNRLG 120
Db 61 RRQIPKARRPREGRTWAQPGVWPLYNNEGCGWAGWLLSPRGSPPSGPTDPRRSNRLG 120

Qy 121 KVIDTLTCGFADLMGYIPLVAGP:GGARALAHGVRLVEDGVNATGNLPCSCFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVAGP:GGARALAHGVRLVEDGVNATGNLPCSCFSIFLLA 180

Qy 181 LLSCLTIPASAYEVNRNSGIVHYVNDSCNSSIVVEAADVIMHTPCGVCPCVEGNSRCWV 240
Db 181 LLSCLTIPASAYEVNRNSGLXVINDCNSISIVVEAADAILHTFGCPCVEEGNASRCWV 240

Qy 241 ALTPTLAARNASVPTTIRRHVDLLVGTAAACSMYVGLDCGSIPLVSQLPFTSPRHET 300
Db 241 AMTPVTATRDGKLPAOTLRHIDILLVGSATLCSALYVGLDCGSLVGLQFTSPSRHWT 300

Qy 301 VQDCNCSYPGHVGSHRAMDMNWSPTTALVVSOLLRIPOAVVDMVAGAKGVLACLA 360
Db 301 TQGCNCSYPGHITGHRWADMNWSPTTALVVAQLLRIPOAILDMTAGAKGVLACLA 360

Qy 361 YYSXVGWNAKYLVALFAGVDGTHTTGRVAGHTTSGFTSLFSSGASQKTLVNTGWS 420
Db 361 YFSMVGNWAKVLVLLFAGVDAETHVTGGSAGHTVSGFVLLAPGAKQNVQLNTGWS 420

Qy 421 HINRTALNCNDSLOTGFPFALFYAHKFNSSGCPERMASCRPIDWFAQGWGITTKPNS 480
Db 421 HINRTALNCNDSLNTGWTAGL:FYHKKFNSSGCPERLASCRPLTDFQGWGPISYANGSGP 480

Qy 481 DQRPYCHYAPPCGVVPASOVCPVYCTPSVTVVGTDRSGVPTYSWGENETDVMILN 540
Db 481 DQRPYCHYPPKPGVGPVPAKSVCPVYCTPSVTVVGTDRSGVPTYSWGENETDVMVFN 540

Qy 541 NTRPPQGNWFCCTWNNSTGFTKTCGPPCNCTGGVNRFTLCPTDCFRKHPATYTKCGSG 600
Db 541 NTRPPLGNWFCCTWNNSTGFTKVCAPPVCGVGGAGNNTLHCTDCFRKHPDATYSCGSG 600

Qy 601 PWTLPCLVDYFYLWHYPCNTLNSIFKRVYVGVGVEHRLNAACNWTGERCNLEDRDRS 660
Db 601 PWTLPCLVDYFYLWHYPCNTYNTIFKRVYVGVGVEHRLNAACNWTGERCNLEDRDRS 660

Qy 661 ELSPLLSTTQWILPCAFPTLIPALSTGLIHLHQNIVDQVLYGVGSFVFAIKWEYIL 720
Db 661 ELSPLLSTTQWILPCAFPTLIPALSTGLIHLHQNIVDQVLYGVGSFVFAIKWEYIV 720

Qy 721 LFLLLADARVACACMLLMLLAQAAALENLVILNAASVAGAGHILSLVFFCAAWYIKG 780
Db 721 LFLLLADARVCSCLWMLLLISQAAALENLVILNAASLAGHILSLVFFCAAWYIKG 780

Qy 781 PLAPGAAAFYGVWVFLILLALLPRAYALDREMAASCGGAVLVGLVFTJLSPYKVELT 840
Db 781 PLAPGAAAFYGVWVFLILLALLPRAYALDREMAASCGGAVLVGLVFTJLSPYKVELT 840

```

```

781 KWVEGAVTTFYGNWVPLLLALLALPQAYALDTEVAASCGVVLVGLMALTLSPYKYIS 840
841 RLJWNLQYFTRAFASMQVWVPLNVRGGRDAIILLTCAVHPELIIFDI TKLLAILGELM 900
841 WCLWNLQYFTRAFASMQVWVPLNVRGGRDAIILLTCAVHPELIIFDI TKLLAILGELM 900
901 VLOAGITRVFVYRAQGL:RACMLVRKVAGGHYVQVFMKLGALTCTYYVNHLPDMDA 960
901 ILQASLLKVPYFVRVQGLLRFCALARKMIGGHYVQVFMKLGALTCTYYVNHLPDMDA 960
961 HAGRLDAVAEPVVSAMETKVIITWGAOTAAAGD:ILGLPVSARGKEIFLGPADSLG 1020
961 HNGRLDAVAEPVVSAMETKVIITWGAOTAAAGD:ILGLPVSARGKEIFLGPADSLG 1020
1021 QGWRLLAPITAYSOQTRGVLCIIITSLTRDKNQVEGEVQVSTATQSFLATCINGVCWT 1080
1021 KGWRLLAPITAYSOQTRGVLCIIITSLTRDKNQVEGEVQVSTATQSFLATCINGVCWT 1080
1081 VYHAGSKTLIAGPKGPIQNTYTVVDLDLVGWQAPPGARSMTPCSCGSSDLYLVTRHADVI 1140
1081 VYHAGSKTLIAGPKGPIQNTYTVVDLDLVGWQAPPGARSMTPCSCGSSDLYLVTRHADVI 1140
1141 PVRRGDSRGLSILSPRVSYLKGSSGGLLPCPSGHVGVFRAAVCTRGVAKAVDFIPVES 1200
1141 PVRRGDSRGLSILSPRVSYLKGSSGGLLPCPSGHVGVFRAAVCTRGVAKAVDFIPVES 1200
1201 METTMSPVFTDNTSPFAVPQTFQVAHLHAPTSGKSTKVPAAAYAGYKVLVLPNSVAA 1260
1201 LETTMSPVFTDNTSSPPVQSFQVAHLHAPTSGKSTKVPAAAYAGYKVLVLPNSVAA 1260
1261 TLGFGAYMSKAGHIDNIRTVGRTITGGSITVSTYTKFLADGCGSGGAYDIIICDECHS 1320
1261 TLGFGAYMSKAGHIDNIRTVGRTITGGSITVSTYTKFLADGCGSGGAYDIIICDECHS 1320
1321 TDSITLIGITVLDQAEATAGARLWLTATPPCSVTVPHENIEEIGLNNNGEIPFYKAI 1380
1321 TDSITLIGITVLDQAEATAGARLWLTATPPCSVTVPHENIEEIGLNNNGEIPFYKAI 1380
1381 PLEIKGGRHLIFCHSKKCDLAALKLGLNAYAYRGLDVSVIPIGDVVVATDAL 1440
1381 PLEIKGGRHLIFCHSKKCDLAALKLGLNAYAYRGLDVSVIPIGDVVVATDAL 1440
1441 MTGFTGPDVIDONTCTVTQVDFSLDPTTITFTTVVQDAVRSQRGRTGGRSRIYR 1500
1441 MTGFTGPDVIDONTCTVTQVDFSLDPTTITFTTVVQDAVRSQRGRTGGRSRIYR 1500
1501 FVTPGEPSCWDFSVLCECYDAGCAWVELT:PAETSVRLRAYLNTPL:PCVQDHLFEWES 1560
1501 FVTPGEPSCWDFSVLCECYDAGCAWVELT:PAETSVRLRAYLNTPL:PCVQDHLFEWES 1560
1561 VFTGLTHIDAHFLSOTKQAGDNFPYLVAYQATVCARAQAPPPSDQMWKCLIRLKPTEL 1620
1561 VFTGLTHIDAHFLSOTKQAGDNFPYLVAYQATVCARAQAPPPSDQMWKCLIRLKPTEL 1620
1621 PTPLLYPLGAVQNEVILTHITIKYIMACMSADLEVT:STWLVGGVLAALAAVCLITGSV 1680
1621 PTPLLYPLGAVQNEVILTHITIKYIMACMSADLEVT:STWLVGGVLAALAAVCLITGSV 1680
1681 VIVGRIITLSCGKPAVPPDREVLYO3PDEMEBCASOLPYIEQGMOLABOFKOKALGLLOTAT 1740
1681 VIVGRIITLSCGKPAVPPDREVLYO3PDEMEBCASOLPYIEQGMOLABOFKOKALGLLOTAT 1740
1741 KQRAAAAPVYESKWRALETFWAKEMNFIISIOYLAGLSTLPGNPAIASMAFTASTSP 1800
1741 KQRAAAAPVYESKWRALETFWAKEMNFIISIOYLAGLSTLPGNPAIASMAFTASTSP 1800
1801 LTTQNTLLFNILGWAQAALAPPSAAGFVQAGTAGAAGSIGLGKVLVDILAGYGAGVA 1860
1801 LTTQNTLLFNILGWAQAALAPPSAAGFVQAGTAGAAGSIGLGKVLVDILAGYGAGVA 1860
1861 GALVAFKVMGSEVPSTEDLVNLLPAIILSPGALVGVVCAAILREYVGPBGAGVOMMRLI 1920
1861 GALVAFKVMGSEVPSTEDLVNLLPAIILSPGALVGVVCAAILREYVGPBGAGVOMMRLI 1920

```

Qy	1921	AFASRGNHVSP	THYVVPESDAAARVTQILSSLTITQULKRLHQWINEDCSPFCSGSWLRDV	1980		
					
Db	1921	AFASRGNHVSP	THYVVPESDAAARVTAILSSLTITQLLRLRQHWISSBCTTPCSCGWLRI	1980		
					
Qy	1981	WDMTCTVLTDF	KTWLQSKLLPRLPKGPVFLSCORGKGVWGDGIMQTCPCGACIAGHYK	2040		
					
Db	1981	WDMTCEVLSD	FKTKLWAKLMPQFPGIPFVSCORGKGVWVDGIMHTRCHCGAEITGHVK	2040		
					
Qy	2041	NGSMRI	VGPRTCSTNMGTFPPINAYITGCTPPAPNYSRALMRVAABEEVETRVDGFH	2100		
Db	2041	NGIMRI	VGPRTRCNMWSSTPPINAYITGCTPPAPNYTALMRVSAABEYVEITRQVGFH	2100		
					
Qy	2101	YVTCMTDNK	PCQVPFAPBFFTEVDGVRHLHRVAPACKPLLRBDVTFOVGINQVLQSQQL	2160		
Db	2101	YVTCMTDNK	PCQVPFSPBFFTELDGVRHLHRFAPPCKPLLRBEVSFVGLHBPVGSQQL	2160		
					
Qy	2161	PCPEP	DVTVLTSMITDPSHITATKRLARGSPPSLASSSASQLSAPSLKATCTTHD	2220		
Db	2161	PCPEP	DVAJVTSMITDPSHITAAAGRRLLARGSPPVASSASQLSAPSLKATCTANH	2220		
					
Qy	2221	SPDADL	TEANILMRQMGNITRVESENKVILDSFEPLHAEGDEREISVAABILRKSK	2280		
Db	2221	SPOAEL	TEANILMRQMGNITRVESENKVILDSPDLVAEEDEREISYPABILRKSR	2280		
					
Qy	2281	PPSALPT	WARPDYNPPLLESKWDPDYPPVVHGCPLPTTAPPIPPRRKRTVVLTESNV	2340		
Db	2281	FAQALP	VWARPDYNPELIVETWKPKDPDYPPVVHGCPLPPKSPVPPPRKRTVVLTESTL	2340		
					
Qy	2341	SSALAE	LAETKTFGSGSSAVDSGTATLPOLASDDGDKGSDVESYSWMPLEGEPPDIL	2400		
Db	2341	STALAE	LAETRSFGSSSTSGITGDNITSSBPAPSGCPDSDAASYSWMPLEGEPPDIL	2400		
					
Qy	2401	SDGWS	TVSBEA-SEDVVCCSMSTVTGALITPCAAESKLPINPLNSLLRHHNMVYAT	2459		
Db	2401	SDGWS	TVSBEANAEDVCCSMSTVTGALVTPCAAEEQKLPINALSNSLLRHHNLVYST	2460		
					
Qy	2460	TSRAS	LROKQVYTRDLQVLDHDPDVVKEMKAKASVTAKLLSIEBACKLTPPHSAKS	2519		
Db	2461	TSRAC	ORQKVTRDLQVLDHSHYQDVLKEYKAAASVKANLLSVEBACSLTPPHSAKS	2520		
					
Qy	2520	FGYGA	OVNLSASAVNHSVWEILDEBTPTDITTIMAKSEVFCVQPEKGRKPPALI	2579		
Db	2521	FGYGA	KVURCHARXAVTHINSVKOLLEDNVPTDITTIMAKNEVFCVQPEKGRKPPALI	2580		
					
Qy	2580	VFDL	GVVCEKMALYDVVSTLPOAVMGSSYFGFYSPQRQVFEIVNTWKSCKPMGFSYD	2639		
Db	2581	VFDL	GVVCEKMALYDVVTKLPVAVMGSSYFGFYSPQQRVFEIVQAWKSKTPMGFSYD	2640		
					
Qy	2640	TRCFD	SVTVTESDI	VERSIVOCCLAPARQAI	BSLTERLYIGPLTNSKQWNGYRRCR	2699
					
Db	2641	TRCFD	SVTVTESDI	ITEBAIYOCCLDLPQARVAIKSLTERLYXVGGPLTNSXGNGCYRRCR	2700	
					
Qy	2700	ASGVLT	ITSCGNLTLCYIKATACRAAKLQDCTMLVNGDDI	AVI	CESAGTQEDAAALRAFT	2759
Db	2701	ASGVLT	ITSCGNLTLCYIKARAACRAAGLQDCTMLVCGDDL	AVVICSAGVQEDAAALRAFT	2760	
					
Qy	2760	EAMTRY	SAPPGDPQ	PEVDLELITSCSNVSVANDASGRKVYITTRDPTPLARAAMETA	2819	
Db	2761	EAMTRY	SAPPGDPQ	PEYDLELITSCSNVSVVAIDGAKRVYITTRDPTPLARAAMETA	2820	
					
Qy	2820	RHTP	NSMLGNI	INVAPTLMARMLMTHFFPSLILAOBLEKALDQCIYGACYSIEPJDLP	2879	
Db	2821	RHTP	VNSWLGNI	INPAFTLMARMLMTHFFSVLTLARDOLEQALDCEIYGACYSIEPJDLP	2880	
					
Qy	2880	QIIE	RLHGLSAFTLHFSY	SPGEINRVASCLRMGLGVPPPLTRWHRARSYVRKLLSOGGGAAT	2939	
Db	2881	PII	QRLHGLSAFSLFSY	SPGEINRVAACLRKLGVPLEAWHRARSYVARLLARGGAAT	2940	
					
Qy	2940	CGRYL	FNWAVTKUL	PTIPASOLDUSGNFVAGYSGGDIYHVSLSRARPRWFPCLLLLS	2999	
Db	2941	CGKYL	FNWAVTKUL	PTIAAAGOLDUSGNWTAGYSGGDIYHVSYSHARPRWFCLLLLA	3000	

```

Qy 3000 VGVGYVLLPNR 3010
    |||||||
Db 3001 AGVGIVLLPNR 3011

RESULT 15
US-09-238-076-20
; Sequence 20, Application US/09238076
; Patent No. US20020102540A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESS: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
US-09-238-076-20

Query Match 88.0%; Score 14090.5; DB 9; Length 3011;
Best Local Similarity 85.3%; Pred.No.0;
Matches 2567; Conservative 217; Mismatches 226; Indels 1; Gaps 1;

Qy 1 MSTNPKPKRKTKNTNRPPQDVKPPGGQIVGGVYLLPRGPRILGVRAKTRKASERSQPRG 60
Db 1 MSTNPKPKRKTKNTNRAPQDVPEFPGGQIVGGVYLLPRGPRILGVRAKTRKTSRSQPRG 60

Qy 61 RRPQIPKARREGRAWAQGYFWPLYGNEGILGWAGWLLSPRGSPSPSGPTDPRRRSRNLG 120
Db 61 RRPQIPKARREGRTWAGPYFWPLYGNEGCGWAGWLLSPRGSPSPSGPTDPRRRSRNLG 120

Qy 121 KVIDLTLCGFADLMGYPLVQAPJGGAARALAHGVRVLEDCVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDLTLCGFADLMGYPLVQAPJGGAARALAHGVRVLEDCVNYATGNLPGCSFSIFLLA 180

Qy 191 LLSCLTTPASAYEVNYSGLIYHVNTDCNSSIVVEADAVIMHTPCVPCVQEGNSSRCWV 240
Db 191 LLSCLTTPASAYQVNSGLIYHVNTDCNSSIVVEADAVIMHTPCVPCVQEGNASRCWV 240

Qy 241 ALTPTTIAARNASVPTTITRRHVDLIVGTAAFCASMYVGDLCGSIFLVGSLFTFSRRHET 300

```


241 AVTPTVATRDGKLPPTQTLRRHIDLLVGSATLCSALYVGDLCGSPVLVQGLTFTEPRRHW 300
 301 VQDCNCSIYPGHVSGHRNANDMMNWSPTTALVUSQLLRIPQAVVDMVAGHWGLAGLA 360
 301 TQSCNCSIYPGHITGHRMANDMMNWSPTAALVVAQLLRIPQAIMDMTAGAWGLAGIA 360
 361 YSMVGNWAKVLIIVALLPAGVDGETHTRGVAGHTTSGFTSLFSSGASOKIQLVNTGWS 420
 361 YFSMVGNWAKVLVLLPAGVDAETHVVGGAHTTAGLVLLTPGAKXONQLINTGWS 420
 421 HINETALNCDSLOTGFPAALFYAHKTNSSCCPERMASCRPIDWFAQGWBITTKPNS 480
 421 HINTALNCDSLSLTGMLAGLFYRHKNSSGCPERLASCRRLTDPAQGWGPIFISANGSG 480
 481 DORPYCHYAPRPGVVPASQVCGPVYCFTEPVPVGTDSGVPYTSWGNENETDMLN 540
 481 DERPYCHYPRPGVPEFAKSVCGPVYCFTEPVPVGTDRSGAPYTSWGANDETDFVLN 540
 541 NTRPPQGNWEGCTWNNSTGFTKTCGPPCNTGGVGNRTLICPTDCFRKHPEATYTKCGS 600
 541 NTRPPLGNWFGCTWNNSTGFTKVCAPFCVIGGVGNRTLICPTDCFRKHPEATYSRCCSG 600
 601 PWLTPRCVLVDYPRLWHPYCTNFSIFKRMVYGVGEHRLNAAONWTRGERCNLEDRDS 660
 601 PWITPRCMVDYPRLWHPYCTNFSIFKRMVYGVGEHRLNAAONWTRGERCNLEDRDS 660
 661 ELSPLLJSTTEWOILPCAFPTLPALSTGLIHLHONIVDQVLYGVSAFVSFAIKWEYIL 720
 661 ELSPLLSTTQWQVPCSFPTLPALSTGLIHLHONIVDQVLYGVSSIASWAIKWEYIV 720
 721 LFLLLADARVACILWMLLIAQAAALENLVLNNAASVAGAHGLSLFLVFFCAAWYIKG 780
 721 LFLLLADARVCSLWMLLISQAAALENLVLNNAASLAGTHGLVSLVFLVFFCAWYILKG 780
 781 RIAPGAAVAFYGVWELLILLALPPRAYALDREMAASCGAVLUGLVLFTLSPIYKVLIT 840
 781 RWVPQAVYAFYGMWFLILLALPPRAYALDREMAASCGAVLUGLVLFTLSPIYKVLIT 840
 841 RLILWLOVFIPEAERHQMVPPLNVRGRDAIILLTCAVHPELIFTKLLALLGLPLM 900
 841 WCMWMLQVFLTRVEAQLHWVPLNVRGRDAIILLTCAVHPELIFTKLLALLGLPLM 900
 901 VLQAGITRVPYVRAQGLIRACMLVRKAGHYVQVFMKLGAALGTYYVYNHLTFLRWA 960
 901 ILQASLLKVPYVRYVQGLIRACMLVRKAGHYVQVFMKLGAALGTYYVYNHLTFLRWA 960
 961 HAGLRDLAVAPVVPVFSAMETKVIITWADTAACGDIILGLPVSARGKIEFLGRADSLG 1020
 961 HNGLRDLAVAPVVPVFSAMETKVIITWADTAACGDIILGLPVSARGKIEFLGRADSLG 1020
 1021 QWRLLAPITAYSQOTRVLGCIITSLTGRDKNOVEGEVQVYSTATQSFATCINGVCWT 1080
 1021 KQWRLLAPITAYAQTRGLLGGIITSLTGRDKNOVEGEVQVYSTATQSFATCINGVCWT 1080
 1081 VTHGASKTLAGKPIITOMYNVNDLVLVGOAPPGARSMTPCSCGSDLYLVTRHADVI 1140
 1081 VYHGAGTRITIASPKGPVIOYNTVNDQLVGVWAPQGSRLTPTCTCGSSDLVTRHADVI 1140
 1141 PVRRGDSRGLISRPVSYLKGSSGGLIPCSCGVYGVFRAAVCTRGVAKAVDITPVES 1200
 1141 PVRRGDSRGLISRPVSYLKGSSGGLIPCSCGVYGVFRAAVCTRGVAKAVDITPVEN 1200
 1201 METTMSPTDNTSTPAPVQFQVLAHLHAPYSGKSTKVPAAAYAGQKVLVLPVSVA 1260
 1201 LETTMSPTDNTSTPAPVQFQVLAHLHAPYSGKSTKVPAAAYAGQKVLVLPVSVA 1260
 1261 TLGFGAYMSKAGHIDPNIETGVRTITGSIYTSYTKFLADGGCGGAYDIICDECHS 1320
 1261 TLGFGAYMSKAGHIDPNIETGVRTITGSIYTSYTKFLADGGCGGAYDIICDECHS 1320
 1321 TOSTTILGTULDOAETAGARVLVATATPPGSSVTVPENIBEIGLSNNGELPFYFKAI 1380
 1321 TDATSIILGTULDOAETAGARVLVATATPPGSSVTVPENIBEIGLSNNGELPFYFKAI 1380

QY 1381 PIRAIKGRHLIFCHSKKCDLAALKTLGLNNAVAYRGLDVSVIPPIGDVVVVATDAL 1440
 Db 1381 PLEVIKGRHLIFCHSKKCDLAALKTLGLNNAVAYRGLDVSIPITNGDVVVVSTDAL 1440
 QY 1441 MTGPTGDFSDVIDCNTCVTVDFPSLDPTFTITTTTVPQDAVSRSQRRGTGRGRSIYR 1500
 Db 1441 MTGPTGDFSDVIDCNTCVTVDFPSLDPTFTITTTTVPQDAVSRSQRRGTGRGRSIYR 1500
 QY 1501 FVTPGEPSPGNFSSVLCVYDAGCAWVELTPEAETSURLRAYLNTPLPYCQDHLPEWES 1560
 Db 1501 FVAPGEPSPGNFSSVLCVYDAGCAWVELTPEAETSURLRAYLNTPLPYCQDHLPEWEG 1560
 QY 1561 VFTGLTHIDAEFFSOTKQAGDNFYLVAQATVCARAQAPPPSWDQWKKCLIRLKPFLHG 1620
 Db 1561 VFTGLTHIDAEFFSOTKQAGDNFYLVAQATVCARAQAPPPSWDQWKKCLIRLKPFLHG 1620
 QY 1621 PTPLLYXLGAVQNVENVILTHPIITKIMACMSADLEVTSTWVLVGVVLAALAAAYCLTGSV 1680
 Db 1621 PTPLLYXLGAVQNVENVILTHPIITKIMACMSADLEVTSTWVLVGVVLAALAAAYCLTGSV 1680
 QY 1681 VIVGRIILSGKPAVVPDREVLYQSFDEMERECASOLPYIEQOMOLAEOPKOKALGLOTAT 1740
 Db 1681 VIVGRIILSGKPAVVPDREVLYQSFDEMERECASOLPYIEQOMOLAEOPKOKALGLOTAT 1740
 QY 1741 KQABAAAPVSVESKRALETFWAKIMNFIISGIIQYLAGLSTLPGNPALASLMAFTASITSP 1800
 Db 1741 RHAEVITPAVQTNWQKLEVPFWAKIMNFIISGIIQYLAGLSTLPGNPALASLMAFTASITSP 1800
 QY 1801 LTTQNTLLFNILGQWAAQOLAPPSSAASAFVGCAGIAGAAVSGIGLKVLVDILAGYGAVA 1860
 Db 1801 LTTQNTLLFNILGQWAAQOLAPPSSAASAFVGCAGIAGAAVSGIGLKVLVDILAGYGAVA 1860
 QY 1861 GALVAFKVMSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGECAVOMNRLI 1920
 Db 1861 GALVAFKVMSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGECAVOMNRLI 1920
 QY 1921 AFASRGNHVSPTHVPESDAAARVTQIISLITITQLKRLHQNEDCSTPCSGSWLRDV 1980
 Db 1921 AFASRGNHVSPTHVPESDAAARVTQIISLITITQLKRLHQNEDCSTPCSGSWLRDI 1980
 QY 1981 WDMITCTVLTDFTKWLQSKLLPRLPGVPLSCQRYKGVWRRGGINGITPCGAQIAGHV 2040
 Db 1981 WDMITCTVLTDFTKWLQSKLLPRLPGVPLSCQRYKGVWRRGGINGITPCGAQIAGHV 2040
 QY 2041 NGSMIIVGPRTCSNTHGTFPIINAYTTCPTSPAPNYSRALWVAABEYVEVTRVGDH 2100
 Db 2041 NGSMIIVGPRTCSNTHGTFPIINAYTTCPTSPAPNYSRALWVAABEYVEVTRVGDH 2100
 QY 2101 YVTGMTDNRKPCQVPAPEFTEVDGVRHLRYAPACKPLLRREDVTFOVGLNQYLVSQ 2160
 Db 2101 YVTGMTDNRKPCQVPAPEFTEVDGVRHLRYAPACKPLLRREDVTFOVGLNQYLVSQ 2160
 QY 2161 PCEBEPDVTULMSLTDPSSHITAEKRIARGSEPSLASSASOLSAKCTTHD 2220
 Db 2161 PCEBEPDVTULMSLTDPSSHITAEKRIARGSEPSLASSASOLSAKCTTHD 2220
 QY 2221 SPDAELIENLWROEMGNITRVESENKVIILDSFEPLHAEGEREISVAEILKSRK 2280
 Db 2221 SPDAELIENLWROEMGNITRVESENKVIILDSFEPLHAEGEREISVAEILKSRK 2280
 QY 2281 PPSALPIWARPDYNPPLLESWKDPDYPVPVHGCPLPTKAPPPPPRKRKTIVLTENV 2340
 Db 2281 FARALPVNARPDYNPPLLESWKDPDYPVPVHGCPLPTKAPPPPPRKRKTIVLTENV 2340
 QY 2341 SSALAEALATKFGSGSSAVDSGTATALPDLASDDGKSDVESYSSMPPLEGEPPDL 2400
 Db 2341 SSALAEALATKFGSGSSAVDSGTATALPDLASDDGKSDVESYSSMPPLEGEPPDL 2400
 QY 2401 SDGWSVTSVESA-SDVYVCCSMTWTGALITPCAAEESKLPINPLSNLLRHHNNVAT 2459
 Db 2401 SDGWSVTSVESA-SDVYVCCSMTWTGALITPCAAEESKLPINPLSNLLRHHNNVAT 2459

2460	QY	TSSASIROKKVTDFDLQVLVDHYRDVLKEMKAKASTVKAALLISTEBEACKLTTPPSAKSK	2519
2461	Db	TSSACQOROKKVTDFDLQVLDSHYQDVLKEVKAASKVKANLLSVBEACSLTPPSSAKSK	2520
2520	QY	FGYGAQDVRLNLSRPANHHISVWEDLLEDETETIDTIMAKSVFVCQPEKGGRKPARLL	2579
2521	Db	FGYGAQDVRRCHAKVAHINSWKKDLLEDSVTFDIIIMAKNEVFCVQPEKGGRKPARLL	2580
2580	QY	VFPDLGVURCEKMAIXDVSTLTPQAVMGSSYGFQYSPKORVEFLVNTWKSKKCPMGFSYD	2639
2581	Db	VFPDLGVURCEKMAIXDVYSKPLAVMGSSYGFQYSPGQREVEFLVQANKSKKTPMGFPYD	2640
2640	QY	TRCPDSTVTRSDIRVRHSIYQCCDLAPEARQATRSITERLYICGPIYNSKGQNCQYRRCR	2699
2641	Db	TRCFDSTVTRSDIRTEBAIYQCCDLDPQARVAIKSLTERLYVGGPITNSRGENCYRRCR	2700
2700	QY	ASGVLITSCNGLITCYLIKATAACRAAKLOCTMLVNGDDLAVTCESAGTOEDAAALRAFT	2759
2701	Db	ASGVLITSCNGLITCYIKAPARRAAGLOCTMLVCGDDLUVICESAGVOEDAAASRAFT	2760
2760	QY	EAMTRYSAPEGDPQPEYDLELITSCSSNSVVAHDASGKRVVYJTRDPTTFLARAAWETA	2819
2761	Db	EAMTRYSAPEGDPQPEYDLELITSCSSNSVVAHDGAGKRVVYJTRDPTTFLARAAWETA	2820
2820	QY	RHTPINSWLGNIIMYAPTLWARMILMTHFFSILLAOQLSKALDCQYGCACIEBIDLDP	2879
2821	Db	RHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQQLANCEITAAACYIEBIDLDP	2880
2880	QY	QIITERLHGLSAFTLHSYSGEINFEVASCLPKIGVPPERTWHRARSVRAKLLSQGGRAAT	2939
2881	Db	PIIQRHLGLSAFTLHSYSGEINFEVASCLPKIGVPPURAKHRARSVRAKLLSGGRAAI	2940
2940	QY	CGRYLFENWAVRTKLIPTIPAAQSOLDISGWFVAGYSGGDIYHISLSIRARPWFPLCLLLS	2999
2941	Db	CGKLYLFENWAVRTKLIPTIPAAQGLDLSGWFITAGYSGGDIYHVSHPARPWAFWFCULLLA	3000
3000	QY	VGVGIVYLLPNR	3010
3001	Db	AGVGIVYLLPNR	3011

Search completed: April 8, 2004, 10:57:23
Job time : 90 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2004, 10:39:25 ; Search time 90 seconds

(without alignments)

9449.644 Million cell updates/sec

Title: US-09-662-454-3

Perfect score: 16009

Sequence: 1 MSTNPKPQKTKRNTNRRPQ.....FPLCLLLSVGVGIYLLPNR 3010

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16009	100.0	3010	2 AAW98022	AAW98022 Infectiou
2	16009	100.0	3010	4 AAB31170	Aab31170 Amino aci
3	15963	99.7	3010	4 AAB59174	Aab59174 Protein e
4	13358	95.9	3010	2 AAR68864	Aar68864 Hepatitis
5	13355	95.9	3010	2 AAR82894	Aar82894 Partial H
6	15332	95.8	3010	2 AAR68622	Aar68622 HCV prote
7	15323	95.7	3010	5 AEG32451	Abg32451 Hepatitis
8	15319	95.7	3010	5 AEG32460	Abg32460 Hepatitis
9	15319	95.7	3010	5 AEG32454	Abg32454 Hepatitis
10	15318	95.7	3010	5 AEG32461	Abg32461 Hepatitis
11	15317	95.7	3010	5 AEG32458	Abg32458 Hepatitis
12	15317	95.7	3010	5 AEG32459	Abg32459 Hepatitis
13	15317	95.7	3010	5 AEG32455	Abg32455 Hepatitis
14	15317	95.7	3010	5 AEG32457	Abg32457 Hepatitis
15	15317	95.7	3010	5 AEG32452	Abg32452 Hepatitis
16	15316	95.7	3010	5 AEG32453	Abg32453 Hepatitis
17	15312.5	95.6	3011	5 AEG32456	Abg32456 Hepatitis
18	15253	95.3	3010	2 AAR20111	Aar20111 Non-A, no
19	15253	95.3	3010	2 AAR20091	Aar20091 Non-A, no
20	15251	95.3	3010	2 AAY06423	Aay06423 Non-A, no
21	15067	94.1	3010	2 AAE30616	Aae30616 Polypepti
22	15064	94.1	3010	5 AAE320477	Aae320477 HCV-S1 fu
23	15041.5	94.0	3011	2 AAR34468	Aar34468 Encoded b
24	15040	93.9	3014	2 AAR54099	Aar54099 NANBHV EI
25	15023	93.8	3010	2 AAR53417	Aar53417 Blood tra

ALIGNMENTS

RESULT 1

AAW98022

ID AAW98022 standard; protein; 3010 AA.

XX AAW98022;

XX 21-JUN-1999 (first entry)

XX Infectious hepatitis C virus genotype 1b strain HC-J4 protein.

XX HCV; infectious clone; infection; diagnosis; therapy; vaccine; screening;
XX assay; antiviral; virucide.

XX Hepatitis C virus.

XX WO9904008-A2.

XX 28-JAN-1999.

XX 16-JUL-1998; 98WO-US014688.

XX 18-JUL-1997; 97US-0053062P.

XX 27-JAN-1998; 98US-00014416.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Yanagi M, Bukh J, Emerson SU, Purcell RH;

XX WPI; 1999-132252/11.

XX N-FSDB; AAX24843.

XX New isolated hepatitis C virus nucleic acids - used to develop products
XX for the diagnosis, prevention and treatment of HCV infections and for
XX developing screening assays.

XX Claim 2; Fig 14G-H; 126pp; English.

XX This protein is encoded by the infectious hepatitis C virus (HCV)
XX genotype 1b strain HC-J4 genome (see AAX24833). HC-J4 was obtained from
XX acute phase plasma of a chimpanzee infected with serum containing HC-
XX J4/91. The infectious nucleic acid sequence can be used to produce
XX chimeric genomes (see AAX24833) consisting of the open reading frames of
XX infectious nucleic acid sequences of other genotypes (including genotypes
XX 1-6) and subtypes (such as 1b, 2a, 2b, 2c, 3a, 4a-f, 5a and 6a) of HCV.
XX The invention also relates to the introduction of mutations or deletions
XX into infectious nucleic acid sequences in order to produce an attenuated
XX HCV virus suitable for vaccine development. Infectious nucleic acid
XX sequences can also be used to produce attenuated virus via passage in

ADD67962 EMCV inte
Aar35207 Hepatitis
Aar34580 Human hep
Aaw98021 Infectiou
Aaw77397 Hepatitis
Abp71460 Amino aci
Aau99289 Hepatitis
Aau61848 HCV H77 C
Aau84597 HCV poly
Aab31169 Amino aci
Aau79221 Hepatitis
Abg73195 MKO-Z vir
Aau98020 Infectiou
Aar30931 Hepatitis
Aaw34480 HCV poly
Aaw40038 HCV poly
Aae22049 Hepatitis
Aar21519 Compiled
Aap59173 Protein e
Add67948 Hepatitis

26 15020.5 93.8 3090 7 ADD67962
27 15014 93.8 3014 2 AAR35207
28 14895 93.0 3010 2 AAR34580
29 14640.5 91.5 3011 2 AAW98021
30 14143.5 88.3 3011 2 AAW77397
31 14143.5 88.3 3011 6 ABP71460
32 14143.5 88.3 3012 5 AAU99289
33 14143.5 88.3 3012 6 AUB61848
34 14131.5 88.3 3011 5 AAU84597
35 14124.5 88.2 3011 4 AAB31169
36 14124.5 88.2 3011 5 AAU79221
37 14124.5 88.2 3011 6 ABG73195
38 14122.5 88.2 3011 2 AAW98020
39 14119.5 88.2 3011 2 AAR30931
40 14119.5 88.2 3011 2 AAW34480
41 14119.5 88.2 3011 2 AAW40038
42 14119.5 88.2 3011 5 AAE22049
43 14110.5 88.1 3011 2 AAR21519
44 14104.5 88.1 3011 4 AAP59173
45 14104.5 88.1 3208 7 ADD67948


```
Qy 1981 WDWTCTVLTDFKTMLOSKLLPRLFGVPLSCQRYKGWVRGDMQMTCPGQAIAGHVK 2040
Db 1981 WDWTCTVLTDFKTMLOSKLLPRLFGVPLSCQRYKGWVRGDMQMTCPGQAIAGHVK 2040
Qy 2041 NGSMRIIVGPRICSNTHGTFPIINAYTTGCPSPAPNYSRALWVAABEYVEVTRVGDFH 2100
Db 2041 NGSMRIIVGPRICSNTHGTFPIINAYTTGCPSPAPNYSRALWVAABEYVEVTRVGDFH 2100
Qy 2101 YVTGMTDNVKPCQVPAPPEFTEVDGVLRLHRYAPACKPLREBTVFQVGLNQYLVGSOL 2160
Db 2101 YVTGMTDNVKPCQVPAPPEFTEVDGVLRLHRYAPACKPLREBTVFQVGLNQYLVGSOL 2160
Qy 2161 PCEPEPDVTLTSMLTDSHITATAKRLARGSPPSLASSASQLSAPSUKACTTHHD 2220
Db 2161 PCEPEPDVTLTSMLTDSHITATAKRLARGSPPSLASSASQLSAPSUKACTTHHD 2220
Qy 2221 SPDADLIEANLLWRQEMGNITRVESENKVILDSFELHAEGEREISVAAEILRSRK 2280
Db 2221 SPDADLIEANLLWRQEMGNITRVESENKVILDSFELHAEGEREISVAAEILRSRK 2280
Qy 2281 PPSALPTIWARDYNPPLLESWKDDYPPVPHGCPPLPTKAPPPRRKKTVVLTSSNV 2340
Db 2281 PPSALPTIWARDYNPPLLESWKDDYPPVPHGCPPLPTKAPPPRRKKTVVLTSSNV 2340
Qy 2341 SSALAEALATKTFGSGSSAVDSGTATALPDLASDDGDKGSDVESYSSMPLEGGPDPL 2400
Db 2341 SSALAEALATKTFGSGSSAVDSGTATALPDLASDDGDKGSDVESYSSMPLEGGPDPL 2400
Qy 2401 SDGWSVTVEEASDDVVCSSSYTWTGALITPCAARESKLPINPLNSLRLHNNVYATT 2460
Db 2401 SDGWSVTVEEASDDVVCSSSYTWTGALITPCAARESKLPINPLNSLRLHNNVYATT 2460
Qy 2461 SRSASLRQKVTFRQLVDLHDYDLVKEMAKASTVKAKLLSTEEACKLTPPHSAKSKF 2520
Db 2461 SRSASLRQKVTFRQLVDLHDYDLVKEMAKASTVKAKLLSTEEACKLTPPHSAKSKF 2520
Qy 2521 GYGAKDVNLSRRVNHRSVWEDLLEDETEPIDTTIMAKSEVFCVQPEKGRPARLIV 2580
Db 2521 GYGAKDVNLSRRVNHRSVWEDLLEDETEPIDTTIMAKSEVFCVQPEKGRPARLIV 2580
Qy 2581 FPDGLGVRCERKALYDVVSTLPQAVMGSSYGFQYSPQRVEFLVNTWKSCKCPMGFSYDT 2640
Db 2581 FPDGLGVRCERKALYDVVSTLPQAVMGSSYGFQYSPQRVEFLVNTWKSCKCPMGFSYDT 2640
Qy 2641 RCFDSTVTESDIRVEESIYQCDLAPBARQAIRSLTERLYTGGBLTNSKGONCGYRRCRA 2700
Db 2641 RCFDSTVTESDIRVEESIYQCDLAPBARQAIRSLTERLYTGGBLTNSKGONCGYRRCRA 2700
Qy 2701 SGVLTTSQCNLTLCYLKATAACRAAKIQDCTMLVNGDDLVIICBSAGTQEDAAALRAFT 2760
Db 2701 SGVLTTSQCNLTLCYLKATAACRAAKIQDCTMLVNGDDLVIICBSAGTQEDAAALRAFT 2760
Qy 2761 AMTRYAPPDGPDPQPEVDLELITSCSNVSVVAHDASGRKRYVYLTRDPTPLARAWEETAR 2820
Db 2761 AMTRYAPPDGPDPQPEVDLELITSCSNVSVVAHDASGRKRYVYLTRDPTPLARAWEETAR 2820
Qy 2821 HTPINSMLGNITIMYAPTLMWAMILMTHTFSTILLAQOLEKALDQIYGACYSIEPLDLPQ 2880
Db 2821 HTPINSMLGNITIMYAPTLMWAMILMTHTFSTILLAQOLEKALDQIYGACYSIEPLDLPQ 2880
Qy 2881 IIERLHGLSAPTLHSYSPGEBINRVASCLRKLGVPPLTRWHRASVRAKLLSQGGRATC 2940
Db 2881 IIERLHGLSAPTLHSYSPGEBINRVASCLRKLGVPPLTRWHRASVRAKLLSQGGRATC 2940
Qy 2941 GRYLFENAVRTKJLKTIPPAASQDLGFWFVAGYSGGDIYHLSZRPRWPPLCLLLSV 3000
Db 2941 GRYLFENAVRTKJLKTIPPAASQDLGFWFVAGYSGGDIYHLSZRPRWPPLCLLLSV 3000
Qy 3001 GVGYLLLPNR 3010
Db 3001 GVGYLLLPNR 3010
```

```
RESULT 2
AAB31170
ID AAB31170 standard; protein; 3010 AA.
XX
XX AAB31170;
XX
XX 02-APR-2001 (first entry)
XX
XX Amino acid sequence of a hepatitis C virus (HCV) clove genotype 1b.
XX
XX Chimeric virus; bovine viral diarrhoea virus; BVDV; hepatitis C virus;
XX HCV; vaccine; viral inhibitor; antiviral.
XX
XX Hepatitis C virus.
XX
XX WO2000075352-A2.
XX
XX 14-DEC-2000.
XX
XX 02-JUN-2000; 2000WO-US015527.
XX
XX 04-JUN-1999; 99US-0137817P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Nam J, Bakh J, Emerson SU, Purcell RH;
XX
XX WPI; 2001-071081/08.
XX
XX N-PSDE; AAC96939.
XX
XX New nucleic acid comprising a chimeric bovine viral diarrhoea virus genome
XX in which the (non-)structural region has been replaced by hepatitis C
XX virus (HCV) genome useful for treating or preventing HCV signs and
XX symptoms.
XX
XX Disclosure; Fig 4G-H; 97pp; English.
XX
XX The specification describes a nucleic acid comprising a chimeric virus
XX genome, specifically bovine viral diarrhoea virus (BVDV) genome in which
XX the (non-)structural region has been replaced by the (non-)structural
XX region of a hepatitis C virus (HCV) genome. The nucleic acids comprising
XX the chimeric virus and the chimeric virus are useful for identifying cell
XX lines capable of supporting the replication of these chimeric viruses, in
XX screening for neutralizing antibodies to HCV of different genotypes, in
XX the production of HCV-BVDV virions, for the development of inactivated or
XX attenuated vaccines to prevent HCV-BVDV in a mammal, in studying the
XX molecular properties of HCV indirectly in vitro, and in identifying
XX inhibitors of viral enzyme activity which would be useful as antiviral
XX agents. Formulations or compositions comprising the chimeric virions may
XX be used to treat or prevent the signs and symptoms of HCV. The present
XX sequence is encoded by a HCV clone, which is used to construct chimeric
XX nucleic acids of the invention
XX
XX Sequence 3010 AA;
XX
XX Query Match 100.0%; Score 16009; DB 4; Length 3010;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 3010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MSINPQQRKTENRRPQDVKFGGGQIVGGVYLLPRRGPRLGVTRATKASERQPRG 60
Db 1 MSINPQQRKTENRRPQDVKFGGGQIVGGVYLLPRRGPRLGVTRATKASERQPRG 60
Qy 61 RRQIPKAREPEGRANAQPGYPWPLYNCEGLWAGWLLSPRSGRPSGWPTDPRRRNLG 120
Db 61 RRQIPKAREPEGRANAQPGYPWPLYNCEGLWAGWLLSPRSGRPSGWPTDPRRRNLG 120
Qy 121 KVDTLTTCGFADLMGYTPIVGAPLGGAAALAHGVLEEDGVNYATGNLPGCCFSIFLLA 180
Db 121 KVDTLTTCGFADLMGYTPIVGAPLGGAAALAHGVLEEDGVNYATGNLPGCCFSIFLLA 180
Qy 181 LLSCLTIPASAYEVRNVSGIYHVHTNDCSSSIYVEAADVIMETPGCVQRCNRRCW 240
```

181	Db		LLSCLTIPASAYEVRNVSGTYHYVTNDSCNSSIYVEAADVLHMPGCVPCVQEGNSRCWV	240
241	Qy		ALTPTLAAANASVPTTTIRHVDLLYCTAAFCASAMVGDLCGSIFLVSQLTFTSPRHET	300
241	Db		AUTPTLAAANASVPTTTIRHVDLLYCTAAFCASAMVGDLCGSIFLVSQLTFTSPRHET	300
301	Qy		VQDCNCSIYPGHVSGHRMAMWMMWNSPTTALVVSQLLRIPQAVDMVAGAHGVLAGLA	360
301	Db		VQDCNCSIYPGHVSGHRMAMWMMWNSPTTALVVSQLLRIPQAVDMVAGAHGVLAGLA	360
361	Qy		YYSWGNWAKVLIVALLFAGVDGETHTTRVAGHTTSGFTSLPSSGASQIKQLVNTNGSW	420
361	Db		YYSWGNWAKVLIVALLFAGVDGETHTTRVAGHTTSGFTSLPSSGASQIKQLVNTNGSW	420
421	Qy		HINRTALNCNDSIQTGFFAALFYAHKFNSSGCPERMASCRPIDWFAOGWGPITYTKPNSS	480
421	Db		HINRTALNCNDSIQTGFFAALFYAHKFNSSGCPERMASCRPIDWFAOGWGPITYTKPNSS	480
481	Qy		DQRPYCHYAPRCGVVPASQVCPVYCTFSPVVVGTTRDSGPVTVSWGENETDVMLN	540
481	Db		DQRPYCHYAPRCGVVPASQVCPVYCTFSPVVVGTTRDSGPVTVSWGENETDVMLN	540
541	Qy		NTRPPQGNFGCTWMSGTGFTKTCGGPPCNIGGVNRTILICPTDCRKHPEANYTKCGSG	600
541	Db		NTRPPQGNFGCTWMSGTGFTKTCGGPPCNIGGVNRTILICPTDCRKHPEANYTKCGSG	600
601	Qy		PWLTPRCLVDYPRYLWHYPCTLNPSIFKVMYVGVVEHRLNAACNMTGRBCNLEDRDS	660
601	Db		PWLTPRCLVDYPRYLWHYPCTLNPSIFKVMYVGVVEHRLNAACNMTGRBCNLEDRDS	660
661	Qy		ELSPLLSTTEWQILPCAFPTTLPALSTGLIHLHQNIVDQYLXGVGSAFVSFAIKWEYIL	720
661	Db		ELSPLLSTTEWQILPCAFPTTLPALSTGLIHLHQNIVDQYLXGVGSAFVSFAIKWEYIL	720
721	Qy		LEPELLADARVCACLWMMLLIAQBAALENLVLNAAASVAGAIGILSFLVFFCAAWYIKG	780
721	Db		LEPELLADARVCACLWMMLLIAQBAALENLVLNAAASVAGAIGILSFLVFFCAAWYIKG	780
781	Qy		RLAPGAAAYFVGWPELILLALLPPRAYALDREMAASCGGAVLVGLVFLTSLSPYKVFELT	840
781	Db		RLAPGAAAYFVGWPELILLALLPPRAYALDREMAASCGGAVLVGLVFLTSLSPYKVFELT	840
841	Qy		RLIWLWLQYFTTRAEAHQVWVPELNVNRRGDDAIIITCAVHPELIFDITKLLAILQLPML	900
841	Db		RLIWLWLQYFTTRAEAHQVWVPELNVNRRGDDAIIITCAVHPELIFDITKLLAILQLPML	900
901	Qy		VLOAGITREYFYFRAOGLIBACMLVRKVAGHYVQVMFKLGALTGYVYVNHITPLRDWA	960
901	Db		VLOAGITREYFYFRAOGLIBACMLVRKVAGHYVQVMFKLGALTGYVYVNHITPLRDWA	960
961	Qy		HAGLRDLAVAVEPVVFSAMETKVIITWGDATTAACGDIILGLPVSARRCKEILFGPADSLEG	1020
961	Db		HAGLRDLAVAVEPVVFSAMETKVIITWGDATTAACGDIILGLPVSARRCKEILFGPADSLEG	1020
1021	Qy		QGWRLIAPITAYSQOTRGVLGCIITSLTRDKNQVEGEVQVWSTATQSFATCINGCWT	1080
1021	Db		QGWRLIAPITAYSQOTRGVLGCIITSLTRDKNQVEGEVQVWSTATQSFATCINGCWT	1080
1081	Qy		VYHAGSKTLIAGPKGPTTQMYTNVDLILVQWAPPGARSMTPCSCGSDLYLVTRHADVI	1140
1081	Db		VYHAGSKTLIAGPKGPTTQMYTNVDLILVQWAPPGARSMTPCSCGSDLYLVTRHADVI	1140
1141	Qy		PVRRGDSRGLSLSPRVSVLYKSGSSGFLIPCPSGHVGVFRAAVCTRGVAKAVDFIPVES	1200
1141	Db		PVRRGDSRGLSLSPRVSVLYKSGSSGFLIPCPSGHVGVFRAAVCTRGVAKAVDFIPVES	1200
1201	Qy		MEITTRSPVFTDNSTPFAVPQTFQVAHLHAPFTGSGKSTKVPAAVAAQYKVLVLPNSVAA	1260
1201	Db		MEITTRSPVFTDNSTPFAVPQTFQVAHLHAPFTGSGKSTKVPAAVAAQYKVLVLPNSVAA	1260
1261	Qy		TLGFGAYMSKAHGIDPNIRTVGRTITTTGGSITYSTYGFADGGCSGGAYDIIITCDECHS	1320

1261	TLGFGAYMSKAHGIDPNIRGCVATIIITGGSIIYTSYVYKPLADGGCGSGAYDIIDIIICDECHS	1320
1321	TDSTIIIGIGTVLDOAETAGARLWLATATPGSVTVPHPNIEEIGLSNNGEIPFYKAI	1380
1321	TDSTIIIGIGTVLDOAETAGARLWLATATPGSVTVPHPNIEEIGLSNNGEIPFYKAI	1380
1381	PTEAIKGRHLIFCHSKKCBELAAKLGLGINAVAYRGLDVSVIPPIGVVVVATDAL	1440
1381	PTEAIKGRHLIFCHSKKCBELAAKLGLGINAVAYRGLDVSVIPPIGVVVVATDAL	1440
1441	MFGTFGDFSDVDCNTCVTQTVDFSDPTFTTETTVQDAVSQRGRTGGRSGIYR	1500
1442	MFGTFGDFSDVDCNTCVTQTVDFSDPTFTTETTVQDAVSQRGRTGGRSGIYR	1500
1501	FVTPGERSGMFDSVLCECYDAGCAWYELTPTAETSVRLAYINTPGLPVQCDHLEPWES	1560
1501	FVTPGERSGMFDSVLCECYDAGCAWYELTPTAETSVRLAYINTPGLPVQCDHLEPWES	1560
1561	VFTGLTHIDAHFLSQTQAGDNPPYIVAYQTVCAPAQAPPSBWMKCLIRLKPILHG	1620
1561	VFTGLTHIDAHFLSQTQAGDNPPYIVAYQTVCAPAQAPPSBWMKCLIRLKPILHG	1620
1621	PTPLLYRLGAVQNSVILITHPTIKYIACMSADLEVVTSTWLVGVLAAALAAAYCLTTGSV	1680
1621	PTPLLYRLGAVQNSVILITHPTIKYIACMSADLEVVTSTWLVGVLAAALAAAYCLTTGSV	1680
1681	VTVGRIILSGKPAVVPDREVLVQEFDEMECASQLPYIBQGMQLAEQFKQKALGLLOTAT	1740
1681	VTVGRIILSGKPAVVPDREVLVQEFDEMECASQLPYIBQGMQLAEQFKQKALGLLOTAT	1740
1741	KQAEAAAPVBSKWRALLETFWAKHMNFISGIQYLAGLSTLPGNPAIASLMAFTASTTSP	1800
1741	KQAEAAAPVBSKWRALLETFWAKHMNFISGIQYLAGLSTLPGNPAIASLMAFTASTTSP	1800
1801	LITQNTILFNILGGWVAQALPSSAASFVGCAGIAAGAAVSGTGLGKVLVDILLAGVGAGVA	1860
1801	LITQNTILFNILGGWVAQALPSSAASFVGCAGIAAGAAVSGTGLGKVLVDILLAGVGAGVA	1860
1861	GALVAFKWSGEVSTEDLNLNLPAILSPCALVGVVCAILRRHVGPGEAVQWNRILI	1920
1861	GALVAFKWSGEVSTEDLNLNLPAILSPCALVGVVCAILRRHVGPGEAVQWNRILI	1920
1921	AFASRGHVSPHYVPSDDAAARVTQLLSLTLITQLLKLHWINEDCSTPCSGSWLBDV	1980
1921	AFASRGHVSPHYVPSDDAAARVTQLLSLTLITQLLKLHWINEDCSTPCSGSWLBDV	1980
1981	NWMICTVLTDFKTMLOSKLIPRGVPPPLSCQGYGKWEGDGIQMTTCPCGAOAGHVK	2040
1981	NWMICTVLTDFKTMLOSKLIPRGVPPPLSCQGYGKWEGDGIQMTTCPCGAOAGHVK	2040
2041	NGSMRIVGPRCSNTWGTPTPINAYTIGPCTPSPAPNYGRALWRVAABEYVEVTVGDFH	2100
2041	NGSMRIVGPRCSNTWGTPTPINAYTIGPCTPSPAPNYGRALWRVAABEYVEVTVGDFH	2100
2101	YVTGMTDNVKPCQVPAPPEFTEVDGVRUHRVAPACKPLLEDVTFQVGLNQYIVGSQL	2160
2101	YVTGMTDNVKPCQVPAPPEFTEVDGVRUHRVAPACKPLLEDVTFQVGLNQYIVGSQL	2160
2161	PECEPDDVTVLTSMLTDPSSHITAEAKRLARGSPSSLASSASQLSAPSLKATCTTHD	2220
2161	PECEPDDVTVLTSMLTDPSSHITAEAKRLARGSPSSLASSASQLSAPSLKATCTTHD	2220
2221	SPDADLIKANLLWQEMGNTTRVSESNKVILDSPEPLHAGDBRETSVAAEILRKSRK	2280
2221	SPDADLIKANLLWQEMGNTTRVSESNKVILDSPEPLHAGDBRETSVAAEILRKSRK	2280
2281	FFSALPIWARPDPYNPPLLESWKOPDYVPPVHVCPLPPTKAPPIPPPRKRKTWVLTESNV	2340
2281	FFSALPIWARPDPYNPPLLESWKOPDYVPPVHVCPLPPTKAPPIPPPRKRKTWVLTESNV	2340
2341	SALABLATKTFGSGSGAVDGTATAPDLADSDGDKGDSVESYSMPPLGEPEGDPDL	2400
2341	SALABLATKTFGSGSGAVDGTATAPDLADSDGDKGDSVESYSMPPLGEPEGDPDL	2400

QY 2401 SDGWSVTSEASEDYVCCSMSTWTGALITPCAABESKLPINPLNSLRHENMYATT 2460
 DB 2401 SDGWSVTSEASEDYVCCSMSTWTGALITPCAABESKLPINPLNSLRHENMYATT 2460
 QY 2461 SRSASLRQKQVTFDRLOVLDDHRYDVLKEMKAKASTVKAKLISIEACKLTPPHSAKSF 2520
 DB 2461 SRSASLRQKQVTFDRLOVLDDHRYDVLKEMKAKASTVKAKLISIEACKLTPPHSAKSF 2520
 QY 2521 GYAKDVRLNLSRAVNHIREWEDLLEDTPTDITIMAKSEVFCVQPEKGGKRPRLIV 2580
 DB 2521 GYAKDVRLNLSRAVNHIREWEDLLEDTPTDITIMAKSEVFCVQPEKGGKRPRLIV 2580
 QY 2581 FPDGLVRVCEKMAKYDVVSTLPAVMSSYGFQYSPKQVREFLVNTWKSCKCPWGSFYDT 2640
 DB 2581 FPDGLVRVCEKMAKYDVVSTLPAVMSSYGFQYSPKQVREFLVNTWKSCKCPWGSFYDT 2640
 QY 2641 RCPDSTVTESDIVERESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKQNCYRRCRA 2700
 DB 2641 RCPDSTVTESDIVERESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKQNCYRRCRA 2700
 QY 2701 SGVLTTSCGNTLCYLKATAACRAAKLQDCTMLVNGDDLVIICESAGTQDAAALRAFT 2760
 DB 2701 SGVLTTSCGNTLCYLKATAACRAAKLQDCTMLVNGDDLVIICESAGTQDAAALRAFT 2760
 QY 2761 AMTRYSAPPDPPQPEYDLELITSCSSNVSAHDASGRVYVLTTRDPTTTLARAAWETAR 2820
 DB 2761 AMTRYSAPPDPPQPEYDLELITSCSSNVSAHDASGRVYVLTTRDPTTTLARAAWETAR 2820
 QY 2821 HTFINSWLGNIIMVAPTLWARMIIMTHFFSILLAQEOLKALDCQIYGACYSIEPLDLQ 2880
 DB 2821 HTFINSWLGNIIMVAPTLWARMIIMTHFFSILLAQEOLKALDCQIYGACYSIEPLDLQ 2880
 QY 2881 IIRLHGLSAFTLHYSYSPGBINRVASCLRKLGVPLPTWHRARSVRAKLLSQGRAATC 2940
 DB 2881 IIRLHGLSAFTLHYSYSPGBINRVASCLRKLGVPLPTWHRARSVRAKLLSQGRAATC 2940
 QY 2941 GRVLFNVAWTKLKTPIPAASQDLDSGWFVAGYSGGDIYHSLSRARPRWFFPLCLLLSV 3000
 DB 2941 GRVLFNVAWTKLKTPIPAASQDLDSGWFVAGYSGGDIYHSLSRARPRWFFPLCLLLSV 3000
 QY 3001 GVGIIYLLPNR 3010
 DB 3001 GVGIIYLLPNR 3010
 RESULT 3
 ID AAB59174 standard; protein; 3010 AA.
 AC AAB59174;
 XX 21-MAR-2001 (first entry)
 DT Protein encoded by infectious Hepatitis C virus 1b genotype.
 DE GBV-B; hepatitis C virus; HCV; vaccine.
 KW Hepatitis C virus.
 OS WC2000075337-A1.
 FN 14-DEC-2000.
 XX 02-JUN-2000; 2000WO-US015293.
 XX 04-JUN-1999; 99US-0137694P.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Bukh J, Yanagi M, Emerson SU, Purcell RH;
 XX WPI; 2001-091214/10.

XX New infectious nucleic acids of the GB virus-B clone, useful for
 PT indirectly studying the molecular properties of hepatitis C virus (HCV)
 PT and in developing vaccines and therapeutics for HCV.
 XX Disclosure; Fig 7; 96pp; English.
 XX The present invention relates to GB virus-B. The nucleic acid molecules
 CC of the invention are useful for indirectly studying the molecular
 CC properties of hepatitis C virus (HCV). The infectious nucleic acid
 CC sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used
 CC in the development of vaccines and therapeutics for HCV
 XX Sequence 3010 AA;
 SQ
 Query Match 99.7%; Score 15963; DB 4; Length 3010;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 3002; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 1 MSTNPKPQKTKRNTNRPPQDVKFGGGOIVGGVLLPRRGLGVRAATKASERSQPRG 60
 DB 1 MSTNPKPQKTKRNTNRPPQDVKFGGGOIVGGVLLPRRGLGVRAATKASERSQPRG 60
 QY 61 RRQPIPKARRPREGRAWAQPGYFWPLYGNEGLGWAGWLLSPGRSPSWGPTDPRRSRNLG 120
 DB 61 RRQPIPKARRPREGRAWAQPGYFWPLYGNEGLGWAGWLLSPGRSPSWGPTDPRRSRNLG 120
 QY 121 KVITDITLTCGFADLMGYIPLVGAPLGGAAARALAHGVRLVLEDGVNYATGNLPGCSFSIFLLA 180
 DB 121 KVITDITLTCGFADLMGYIPLVGAPLGGAAARALAHGVRLVLEDGVNYATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTIPASAVEVRNVSGIYHVITNDCSNSSIVIEAADVIMHTPGCVPCQVQGNSSRCWV 240
 DB 181 LLSCLTIPASAVEVRNVSGIYHVITNDCSNSSIVIEAADVIMHTPGCVPCQVQGNSSRCWV 240
 QY 241 ALTPTLAANASVPTTIRRHVDLLVGTAAAFCSAMTVGLCGSIFLVSOLFSPSRHET 300
 DB 241 ALTPTLAANASVPTTIRRHVDLLVGTAAAFCSAMTVGLCGSIFLVSOLFSPSRHET 300
 QY 301 VQDCNCSIYPGHVSGHRMADMMNWSPTTALVVSOLLRIPOAVDMVAGAHVGLAGLA 360
 DB 301 VQDCNCSIYPGHVSGHRMADMMNWSPTTALVVSOLLRIPOAVDMVAGAHVGLAGLA 360
 QY 361 YYSVMGNWAKLIVALLFAGVDGETHTTGRVAGHTTSGFTSLFSSGASOKIOLVNTGWS 420
 DB 361 YYSVMGNWAKLIVALLFAGVDGETHTTGRVAGHTTSGFTSLFSSGASOKIOLVNTGWS 420
 QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCPIDWFAQGWGPITYTKPNSS 480
 DB 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCPERMASCPIDWFAQGWGPITYTKPNSS 480
 QY 481 DQRPYCHVYAPRPGVVPASOVCPVYCTFPPSVVVGTTTDRSGVPTYSWGENETDVMLLN 540
 DB 481 DQRPYCHVYAPRPGVVPASOVCPVYCTFPPSVVVGTTTDRSGVPTYSWGENETDVMLLN 540
 QY 541 NTRPPQGNWFCGTWMSSTGTTCGGPPPCNIGGVGNRTLCPTDCPRKPEATYTKGSG 600
 DB 541 NTRPPQGNWFCGTWMSSTGTTCGGPPPCNIGGVGNRTLCPTDCPRKPEATYTKGSG 600
 QY 601 PWLTPRCLVDYPIYRLWHYPCTIANFSIPKVMYVGGVEHRLNAACNWTGRGRCNLED RDS 660
 DB 601 PWLTPRCLVDYPIYRLWHYPCTIANFSIPKVMYVGGVEHRLNAACNWTGRGRCNLED RDS 660
 QY 661 ELSPLILSTTWEQILLPCAFITTLPALSTGLJHLHQNIVDVQYLYGVGSFAFVSPAIKWEYIL 720
 DB 661 ELSPLILSTTWEQILLPCAFITTLPALSTGLJHLHQNIVDVQYLYGVGSFAFVSPAIKWEYIL 720
 QY 721 LLFLLADARVCACLMWMLLIAQAEALENLVINAAVAGAHGILSLFVFFCAAWIKG 780
 DB 721 LLFLLADARVCACLMWMLLIAQAEALENLVINAAVAGAHGILSLFVFFCAAWIKG 780
 QY 781 RLAPGAAYAFYGVWPLILLIALLPRAVALDRENAASCGGAVLVGLVFLTLSPYKVFELT 840
 DB 781 RLAPGAAYAFYGVWPLILLIALLPRAVALDRENAASCGGAVLVGLVFLTLSPYKVFELT 840

781 RLAPGAAYAFYGVWPELLLLLLLPPRAYALDREMAASCQGAVALVGLVFLTLSPYKVFLLT 840
841 RLILWKLQYFTRABAHQVWVPLNVGGRDAIILLCAVHPELIFDITKLLAILGLPLM 900
841 RLILWKLQYFTRABAHQVWVPLNVGGRDAIILLCAVHPELIFDITKLLAILGLPLM 900
901 VLQAGITRVPYFVBAQGLIRACMLVRKVGAGHYQVMFKLGAITGYVYNHLTPLRDWA 960
901 VLQAGITRVPYFVBAQGLIRACMLVRKVGAGHYQVMFKLGAITGYVYNHLTPLRDWA 960
961 HAGLRDLAVAPVVPFVSAMETKVTWGDADPAACODIILGLPVSARKEKFIPLGADSLG 1020
961 HAGLRDLAVAPVVPFVSAMETKVTWGDADPAACODIILGLPVSARKEKFIPLGADSLG 1020
1021 QGWELLAPITAYSQOTRGVLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCINGVCWT 1080
1021 QGWELLAPITAYSQOTRGVLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCINGVCWT 1080
1081 VYHAGSKTLAGPKGPIQMYTNVDLILVGMQAPPGARSMTPCSCGSSDLYLVTRHADVI 1140
1081 VYHAGSKTLAGPKGPIQMYTNVDLILVGMQAPPGARSMTPCSCGSSDLYLVTRHADVI 1140
1141 PVRRGDSRGSLLSPRVSYLKGSGGPLLCPSGHVGVFRFAAVCTRGVAKAVDFIPVES 1200
1141 PVRRGDSRGSLLSPRVSYLKGSGGPLLCPSGHVGVFRFAAVCTRGVAKAVDFIPVES 1200
1201 METTMRSPVFTDNSTPPAVPQTFQVAHLHAPTGSKSTKVEFAAQAQYKVLVLPNSVAA 1260
1201 METTMRSPVFTDNSTPPAVPQTFQVAHLHAPTGSKSTKVEFAAQAQYKVLVLPNSVAA 1260
1261 TLGFGAYMSKAHGDIPNIRTOVRNITTTGSGTISTYTKGFLADGGCGSGAYDIIICDECHS 1320
1261 TLGFGAYMSKAHGDIPNIRTOVRNITTTGSGTISTYTKGFLADGGCGSGAYDIIICDECHS 1320
1321 TDSITILGIGHVLQDAETAGARLVVLAATPPGVSIVVPHNIEBIGLSNNGEIEFYKAI 1380
1321 TDSITILGIGHVLQDAETAGARLVVLAATPPGVSIVVPHNIEBIGLSNNGEIEFYKAI 1380
1381 PIEAIKGRHLIFCHSKKKCDLAAKTGLGINAVAYYRGLDVSVIPIIGDVVVVATDAL 1440
1381 PIEAIKGRHLIFCHSKKKCDLAAKTGLGINAVAYYRGLDVSVIPIIGDVVVVATDAL 1440
1441 MTGFTGPDSDVIDCNTQVTVDFSLDPTFTIETTVQDAVRSQRGRGSGIYR 1500
1441 MTGFTGPDSDVIDCNTQVTVDFSLDPTFTIETTVQDAVRSQRGRGSGIYR 1500
1501 FVTPGERPSGMFDSVLCBCEYDAGCAWYELTPAETSVELRAYLNTPLGVQDHLFEWES 1560
1501 FVTPGERPSGMFDSVLCBCEYDAGCAWYELTPAETSVELRAYLNTPLGVQDHLFEWES 1560
1561 VFTGLTHIDAHFLSQTQAGDNFYLVAQYATVCARAQAPPPSDQMWKCLIRLKPILHG 1620
1561 VFTGLTHIDAHFLSQTQAGDNFYLVAQYATVCARAQAPPPSDQMWKCLIRLKPILHG 1620
1621 PTPLLYRLGAVQNEVILTHPIITKYIMACNSADLEVVTTSTWLVGVLAAALAYCCTTGSV 1680
1621 PTPLLYRLGAVQNEVILTHPIITKYIMACNSADLEVVTTSTWLVGVLAAALAYCCTTGSV 1680
1681 VIVGRILISGXPAVVPDREVLVQFDEMEECASQLPYIEQGMQLAEQFKQKALGLLOTAT 1740
1681 VIVGRILISGXPAVVPDREVLVQFDEMEECASQLPYIEQGMQLAEQFKQKALGLLOTAT 1740
1741 KOAEAAAPVWBSKPALETFWAKHWNFIISGQYLAGLSTLPGNPATIASLMAFTASTISP 1800
1741 KOAEAAAPVWBSKPALETFWAKHWNFIISGQYLAGLSTLPGNPATIASLMAFTASTISP 1800
1801 LTTQNTLILGNLGGVAAQALAPPSSAASAFVAGIAGAAVGSIGLKVAVDILLAGVGAGA 1860
1801 LTTQNTLILGNLGGVAAQALAPPSSAASAFVAGIAGAAVGSIGLKVAVDILLAGVGAGA 1860
1861 GALVAFKMSGEVPTEDLVNLLPAILSPGALVGVWCVAAILRRHVGPGEAGVQWNNRLI 1920
1861 GALVAFKMSGEVPTEDLVNLLPAILSPGALVGVWCVAAILRRHVGPGEAGVQWNNRLI 1920

1921 AFASRGNHVSPTHYVPESDAAARVTQILSSLLITQILKELHQMINEDECSTPCSGSWLRDV 1980
1921 AFASRGNHVSPTHYVPESDAAARVTQILSSLLITQILKELHQMINEDECSTPCSGSWLRDV 1980
1981 WDWICVILDFKTLWLOSULLPRLPGVFFLSCORGKYGVRGDIQMTTCPCGAQIAGHYK 2040
1981 WDWICVILDFKTLWLOSULLPRLPGVFFLSCORGKYGVRGDIQMTTCPCGAQIAGHYK 2040
2041 NGSMRIVGPRTCNSNTWEGTFFPINAYITGCTPSPADNYSRALWRVAAEYVEVTRVGDFH 2100
2041 NGSMRIVGPRTCNSNTWEGTFFPINAYITGCTPSPADNYSRALWRVAAEYVEVTRVGDFH 2100
2101 YVYGMTTDNVKPCQVPAPEPTEVDGVLHRYAPACKPLLREDVTTFQVGLAQYLVGSOL 2160
2101 YVYGMTTDNVKPCQVPAPEPTEVDGVLHRYAPACKPLLREDVTTFQVGLAQYLVGSOL 2160
2161 PCPEPDVTLTSMITDPSHITAEATAKRRLARGSPPLSSASSASQLSAPSLKATCTTHD 2220
2161 PCPEPDVTLTSMITDPSHITAEATAKRRLARGSPPLSSASSASQLSAPSLKATCTTHD 2220
2221 SPDADLIEANLLWRQWGMGNITRVESENKVILDSPEPLHABGDEREISVAABILKRSK 2280
2221 SPDADLIEANLLWRQWGMGNITRVESENKVILDSPEPLHABGDEREISVAABILKRSK 2280
2281 FPSALPIWAPDYNPPLLESWKDPDYVPPVHGCPLPPTKAPPIPPPRKRVVLTESNV 2340
2281 FPSALPIWAPDYNPPLLESWKDPDYVPPVHGCPLPPTKAPPIPPPRKRVVLTESNV 2340
2341 SSALAEIATKTFCSSSSAVDSCTNTALPDLASDDGDKGSDVESYSSMPLEGECDPL 2400
2341 SSALAEIATKTFCSSSSAVDSCTNTALPDLASDDGDKGSDVESYSSMPLEGECDPL 2400
2401 SDGSWSIVSEASEDVVCCSMSTWTGALITPCAAEESKLPINPLNSILRHHNMYATT 2460
2401 SDGSWSIVSEASEDVVCCSMSTWTGALITPCAAEESKLPINPLNSILRHHNMYATT 2460
2461 SRASLRQKXVTPDRILQVLDHVRDVLKEMKAKASTVAKLILSIEACKLTPPHSKSKF 2520
2461 SRASLRQKXVTPDRILQVLDHVRDVLKEMKAKASTVAKLILSIEACKLTPPHSKSKF 2520
2521 GYKAKVRNLSRAVNHISRWEDLLEDETPTDITIMAKSEVFCVQPEKGGKBPRLIV 2580
2521 GYKAKVRNLSRAVNHISRWEDLLEDETPTDITIMAKSEVFCVQPEKGGKBPRLIV 2580
2581 FPDGLVRVCEKVALYDVVSTLPQAVMGSSYGFQYSPKQVVEFLVNIWKSCKCPMGFSYDT 2640
2581 FPDGLVRVCEKVALYDVVSTLPQAVMGSSYGFQYSPKQVVEFLVNIWKSCKCPMGFSYDT 2640
2641 RCFDSTVTESDIRVBSIYQCCDLAPARQAIHSLTERLYIGGLTNSKQNGYRRCRA 2700
2641 RCFDSTVTESDIRVBSIYQCCDLAPARQAIHSLTERLYIGGLTNSKQNGYRRCRA 2700
2701 SGVLTTCSCGNLTICYLKATAACRAAKLQDCTMLVNGDDLVI CESAGTOEDAAALRAFT 2760
2701 SGVLTTCSCGNLTICYLKATAACRAAKLQDCTMLVNGDDLVI CESAGTOEDAAALRAFT 2760
2761 AMTRYAPPDPPQPEYDLELITSCSSNVSVAHDAAGKRVYIYTRDPTTPLAARAAETAR 2820
2761 AMTRYAPPDPPQPEYDLELITSCSSNVSVAHDAAGKRVYIYTRDPTTPLAARAAETAR 2820
2821 HTPINSHLGNITMVAPTLWARMILMHFFSILLAQOLEKALDCQYGYCYSIEPDLDPQ 2880
2821 HTPINSHLGNITMVAPTLWARMILMHFFSILLAQOLEKALDCQYGYCYSIEPDLDPQ 2880
2881 IIEELHGLSAFTLHYSYSGPEINEVASCLRLKGVPPPLRTWHRARSVRAKLLSQGGAATC 2940
2881 IIEELHGLSAFTLHYSYSGPEINEVASCLRLKGVPPPLRTWHRARSVRAKLLSQGGAATC 2940
2941 GRVLFNNAVATKILTPIPAASQDLSGWFVAGYSGGDIYHESLRAPRPFPLCLLLLSV 3000
2941 GRVLFNNAVATKILTPIPAASQDLSGWFVAGYSGGDIYHESLRAPRPFPLCLLLLSV 3000

Db 781 RLVPGBAALYGVWELLJLLALLPBRAYAMDREMAASCGBAVFVGLVLLTLSPYKVFLLA 840
Qy 841 RLJWLQVPI TRAEAHQWVWVPLNVRGRDAIILLTCVAVHPELIPDTIKLILAILGLPM 900
Db 841 RLJWLQVPI TRAEAHQWVWVPLNVRGRDAIILLTCVAVHPELIPDTIKLILAILGLPM 900
Qy 901 VLOAGITRVPYVRAQGLIRACWLKVKAGGHVQVWFMKLGALGTYYVNHLTPLRDWA 960
Db 901 VLOAGITRVPYVRAQGLIRACWLKVKAGGHVQVWFMKLGALGTYYVNHLTPLRDWA 960
Qy 961 HAGLRDLAVAVFPVPSMETKVIITWADTAACGDIILGLPVARSARKBIFILGPADSLEG 1020
Db 961 HAGLRDLAVAVFPVPSMETKVIITWADTAACGDIILGLPVARSARKBIFILGPADSLEG 1020
Qy 1021 QWRLLAPITAYSOQTRGVLCIITSLGRDNKNOVEGEVQVSTATOSFLATCINGVCWT 1080
Db 1021 QWRLLAPITAYSOQTRGVLCIITSLGRDNKNOVEGEVQVSTATOSFLATCINGVCWT 1080
Qy 1081 VHGAGSKTLAGPKGPIITQMTYNVDLDLVGWOAPPGARSMTPCSCSSDLIYVTRHADVI 1140
Db 1081 VHGAGSKTLAGPKGPIITQMTYNVDLDLVGWOAPPGARSMTPCSCSSDLIYVTRHADVI 1140
Qy 1141 PVRRGDSGSLSPRPVSYLKGSSGGPILCPGSHVWGVFRAAVCTRGVAKAVDFIPVES 1200
Db 1141 PVRRGDSGSLSPRPVSYLKGSSGGPILCPGSHVWGVFRAAVCTRGVAKAVDFIPVES 1200
Qy 1201 METTMSPVFTDNSTPPAVPOFQVAHLHAPTGSKSTKVPRAAYAAQGVKTVLINESVAA 1260
Db 1201 METTMSPVFTDNSTPPAVPOFQVAHLHAPTGSKSTKVPRAAYAAQGVKTVLINESVAA 1260
Qy 1261 TLGFGAYMSKAHGDINIRTVGTITGGSIITYSGKFIADGCGSGGAYDIIICDECHS 1320
Db 1261 TLGFGAYMSKAHGDINIRTVGTITGGSIITYSGKFIADGCGSGGAYDIIICDECHS 1320
Qy 1321 TDSITTLIGTVDQABTAGARLVILATATPPQSVTVPHNIBEIGLSNNGHIPPFGKAI 1380
Db 1321 TDSITTLIGTVDQABTAGARLVILATATPPQSVTVPHNIBEIGLSNNGHIPPFGKAI 1380
Qy 1381 PIRAIKGRHLIFCHSKKKDELAALKITGLNNAVAYGLDVSVPPIGDVVVATDAL 1440
Db 1381 PIRAIKGRHLIFCHSKKKDELAALKITGLNNAVAYGLDVSVPPIGDVVVATDAL 1440
Qy 1441 MTGFTGDFSDVDCNTCVTVDFSLDPTFTITTTVPQDAVRSQRGRGSGIYR 1500
Db 1441 MTGFTGDFSDVDCNTCVTVDFSLDPTFTITTTVPQDAVRSQRGRGSGIYR 1500
Qy 1501 FVTPGERSGMPDSSVLCBQYDAGCAWELTPAETSURLRAYLNTPLPYCQDHLFPWES 1560
Db 1501 FVTPGERSGMPDSSVLCBQYDAGCAWELTPAETSURLRAYLNTPLPYCQDHLFPWES 1560
Qy 1561 VFTGLTHIDAEFLSOTKQAGDNPPYLVAQATVCARAQAPPSPWDQWKKLIRKPTLHG 1620
Db 1561 VFTGLTHIDAEFLSOTKQAGDNPPYLVAQATVCARAQAPPSPWDQWKKLIRKPTLHG 1620
Qy 1621 PTPLLYELGAVQNEVILTHPTIKYIMACMSADLEWVTSTWVLVGVLAALAAAYCLITGVS 1680
Db 1621 PTPLLYELGAVQNEVILTHPTIKYIMACMSADLEWVTSTWVLVGVLAALAAAYCLITGVS 1680
Qy 1681 VIVGRIIILSGKPAVPREVILYQEFDEMEBCASOLPYIEQWOLABOFKOKALGLLOTAT 1740
Db 1681 VIVGRIIILSGKPAVPREVILYQEFDEMEBCASOLPYIEQWOLABOFKOKALGLLOTAT 1740
Qy 1741 KOAEAAAPVVESKWRALLETWAKHWNFIISQIYLAGLSTLPGNPAIASMAFTASITSP 1800
Db 1741 KOAEAAAPVVESKWRALLETWAKHWNFIISQIYLAGLSTLPGNPAIASMAFTASITSP 1800
Qy 1801 LTTQNTLLFNILGQWAAQALAPPAAAFVAGIAGAAVGSIGIGKVLVDILAGYGAVGA 1860
Db 1801 LTTQNTLLFNILGQWAAQALAPPAAAFVAGIAGAAVGSIGIGKVLVDILAGYGAVGA 1860
Qy 1861 GALVAFKVMGSEVSTEDLVNLLFAILSPGALVGVVCAAIIRRHVGPGBGAVQWNRLLI 1920

Db 1861 GALVAFKVMGSEVSTEDLVNLLFAILSPGALVGVVCAAIIRRHVGPGBGAVQWNRLLI 1920
Qy 1921 AFASRGHVSPTHVVPESDAAARVTQILSSITITQALLKRLHQWINEDECSTPCSSWLKDV 1980
Db 1921 AFASRGHVSPTHVVPESDAAARVTQILSSITITQALLKRLHQWINEDECSTPCSSWLKDV 1980
Qy 1981 WDNICTVLDFKTLQSKLLPRLPGVFPFLSCQRYKGVWRGDGIMQTTCPGAGIAGHYK 2040
Db 1981 WDNICTVLDFKTLQSKLLPRLPGVFPFLSCQRYKGVWRGDGIMQTTCPGAGIAGHYK 2040
Qy 2041 NGSMRIIVGPRCSNTHGTTRPINAATGCTPSPAPNYSRALWVAAAEVVEVTRVGDFFH 2100
Db 2041 NGSMRIIVGPRCSNTHGTTRPINAATGCTPSPAPNYSRALWVAAAEVVEVTRVGDFFH 2100
Qy 2101 YVTGMTDNVKKPCQCPVPAPEFFTEVDGVRHRYAPACKPLREDVTVQVGLNQLVVSQ 2160
Db 2101 YVTGMTDNVKKPCQCPVPAPEFFTEVDGVRHRYAPACKPLREDVTVQVGLNQLVVSQ 2160
Qy 2161 PCEBPDVTVLTSMLTDPSSHITAEKRLARGSPPLASSSSASQLSAPSLKATCTTHHD 2220
Db 2161 PCEBPDVTVLTSMLTDPSSHITAEKRLARGSPPLASSSSASQLSAPSLKATCTTHHD 2220
Qy 2221 SPDADLIEANLLWROEMGCGNITRVESENKVVILDSFEPHAEGBERBISVAAEILRKSRK 2280
Db 2221 SPDADLIEANLLWROEMGCGNITRVESENKVVILDSFEPHAEGBERBISVAAEILRKSRK 2280
Qy 2281 FFSALPIWARPDYNPPLLESWKDPDYVPVHVGCPPLPTKAPPIPPPRKKTIVLTESNV 2340
Db 2281 FFSALPIWARPDYNPPLLESWKDPDYVPVHVGCPPLPTKAPPIPPPRKKTIVLTESNV 2340
Qy 2341 SSALAEATKTFGSSGSSAVDSGTATLPDASDDGKGSVSESVMPPLEGEPDPL 2400
Db 2341 SSALAEATKTFGSSGSSAVDSGTATLPDASDDGKGSVSESVMPPLEGEPDPL 2400
Qy 2401 SDGSMSTVSEASEDWCSSMSYWTGALITPCAAESKLPINPLSNLLRHHNNVYAT 2460
Db 2401 SDGSMSTVSEASEDWCSSMSYWTGALITPCAAESKLPINPLSNLLRHHNNVYAT 2460
Qy 2461 SRSASLRQKKVTFDLQVLDHYRDVLKEMKAKASTVKARLLSIEEACKLTPPHSAKSKF 2520
Db 2461 SRSASLRQKKVTFDLQVLDHYRDVLKEMKAKASTVKARLLSIEEACKLTPPHSAKSKF 2520
Qy 2521 GYGAKDVRMLSRVNHRSVWEDLEDETETIDITAKSEVFCVQPEKGEKPARLIV 2580
Db 2521 GYGAKDVRMLSRVNHRSVWEDLEDETETIDITAKSEVFCVQPEKGEKPARLIV 2580
Qy 2581 FPDGLGVCEKMALYDVVSTLTPQAVMGSGYQVSPKQFVEFLVNTWKSCKCPMGFSYDT 2640
Db 2581 FPDGLGVCEKMALYDVVSTLTPQAVMGSGYQVSPKQFVEFLVNTWKSCKCPMGFSYDT 2640
Qy 2641 RCFDSTVTSDDRVEESIYQCCDLAPAPARQATRSITERLYIGGLTNSKGQNGYRCEA 2700
Db 2641 RCFDSTVTSDDRVEESIYQCCDLAPAPARQATRSITERLYIGGLTNSKGQNGYRCEA 2700
Qy 2701 SGVLTSSCGNTLTCYLKATAACRAAKLOCTMVLNGDDLAVICESAGTODEAALRAFTE 2760
Db 2701 SGVLTSSCGNTLTCYLKATAACRAAKLOCTMVLNGDDLAVICESAGTODEAALRAFTE 2760
Qy 2761 AMTRYAPPDPPQPPYDLELITSCSSNVSVAFDASGKEVYVLTTRDPTTTPPLARAAWETAR 2820
Db 2761 AMTRYAPPDPPQPPYDLELITSCSSNVSVAFDASGKEVYVLTTRDPTTTPPLARAAWETAR 2820
Qy 2821 HTPINSWLGNIIMYAPTLWARMILMTHFFSIIIAQOEKALDCQYIGACYIEPLDLPO 2880
Db 2821 HTPINSWLGNIIMYAPTLWARMILMTHFFSIIIAQOEKALDCQYIGACYIEPLDLPO 2880
Qy 2881 IIRLHGLSAFTLHSSYSPGEINPVSCLFKGVPPURTWRHARSVRAKLLSOGGAAATC 2940
Db 2881 IIRLHGLSAFTLHSSYSPGEINPVSCLFKGVPPURTWRHARSVRAKLLSOGGAAATC 2940
Qy 2941 GRYLFENWAVTKLKTPIPAASQDLISGMFVAGYSGGDIYHLSLSRARPRWFFLCLLLSV 3000
Db 2941 GRYLFENWAVTKLKTPIPAASQDLISGMFVAGYSGGDIYHLSLSRARPRWFFLCLLLSV 3000

QY 1261 TLGFGYMSKAGIDNIRTYGTHITGSGITTYSTYKFLADGCGSGGVADIIICDECHS 1320
 Db 1261 TLGFGYMSKAGIDNIRTYGTHITGSGITTYSTYKFLADGCGSGGVADIIICDECHS 1320
 QY 1321 TDSTTILGIGTVDQAGTARLWVLATATPPGSGVTVPHENIEEIGLSNNGEIPFYGKAI 1380
 Db 1321 TDSTTILGIGTVDQAGTARLWVLATATPPGSGVTVPHENIEEIGLSNNGEIPFYGKAI 1380
 QY 1381 PISAIGKGRHLIFCHSKKKODELAALKTGLGKNAVAYRGLDVSVTPPGVGVVATDAL 1440
 Db 1381 PISAIGKGRHLIFCHSKKKODELAALKTGLGKNAVAYRGLDVSVTPPGVGVVATDAL 1440
 QY 1441 MTGFTGDFDSVIDCNTCTVTQVDFSLDPTFTTITVPOAVRSORRGRTGRSGIYR 1500
 Db 1441 MTGFTGDFDSVIDCNTCTVTQVDFSLDPTFTTITVPOAVRSORRGRTGRSGIYR 1500
 QY 1501 FVTGFRPSGMFSSVLCYDAGCAWYELTPAETSVRLAYNTPLPVQCDELFWES 1560
 Db 1501 FVTGFRPSGMFSSVLCYDAGCAWYELTPAETSVRLAYNTPLPVQCDELFWES 1560
 QY 1561 VFTGLTHIDAFPLSOTKQAGDNPPYLVAOATVCARACAPPSQDQWKCLIRKPTLHG 1620
 Db 1561 VFTGLTHIDAFPLSOTKQAGDNPPYLVAOATVCARACAPPSQDQWKCLIRKPTLHG 1620
 QY 1621 PTLLEYELGAVQNEVILTHPTIKYIMACMSADLEVVTSTWLVGGVLAALAAVCLITGVS 1680
 Db 1621 PTLLEYELGAVQNEVILTHPTIKYIMACMSADLEVVTSTWLVGGVLAALAAVCLITGVS 1680
 QY 1681 VIVGRIILSKPAPVPREVLYQBFDEMBECASQIPYIEQGMQLABOPKOKALGLIQTAT 1740
 Db 1681 VIVGRIILSKPAPVPREVLYQBFDEMBECASQIPYIEQGMQLABOPKOKALGLIQTAT 1740
 QY 1741 KQAEAAAPVVEKRWALETFWAKHWNFIISGIVLAGLSTLPGNPATASLMAFTASITSP 1800
 Db 1741 KQAEAAAPVVEKRWALETFWAKHWNFIISGIVLAGLSTLPGNPATASLMAFTASITSP 1800
 QY 1801 LTTQNTLLFNILGGWAAQAAPPASAFAVYAGIAGAAVSGISGLGKVLVHILAGYGAGVA 1860
 Db 1801 LTTQNTLLFNILGGWAAQAAPPASAFAVYAGIAGAAVSGISGLGKVLVHILAGYGAGVA 1860
 QY 1861 GALVAFKVMGSEVSTEDLNLPAILSPGALVGVVCAAILRHRVOPGEGAVQWNRLLI 1920
 Db 1861 GALVAFKVMGSEVSTEDLNLPAILSPGALVGVVCAAILRHRVOPGEGAVQWNRLLI 1920
 QY 1921 AFASRGNHVSPTHVPSDAAARVTQILSSITITQLLKRHLQWINEDECSTPCSGSWLRDV 1980
 Db 1921 AFASRGNHVSPTHVPSDAAARVTQILSSITITQLLKRHLQWINEDECSTPCSGSWLRDV 1980
 QY 1981 WDWICTVLTDFTKWLQSKLLPRLPGVPLSCORGKYGVRGDIQMTTCPOGAQIAGHVK 2040
 Db 1981 WDWICTVLTDFTKWLQSKLLPRLPGVPLSCORGKYGVRGDIQMTTCPOGAQIAGHVK 2040
 QY 2041 NGSRIIVGPRTCSTNTHGTFPINAATTCPTSPAPNYSRALRVAAEVEVTVRGDFH 2100
 Db 2041 NGSRIIVGPRTCSTNTHGTFPINAATTCPTSPAPNYSRALRVAAEVEVTVRGDFH 2100
 QY 2101 YVTGMTTNDVNCPCQVPAPEFTEVDGVRLLRYAPACKPLLRDVTFOVGLNQVLVGSQ 2160
 Db 2101 YVTGMTTNDVNCPCQVPAPEFTEVDGVRLLRYAPACKPLLRDVTFOVGLNQVLVGSQ 2160
 QY 2161 PCEPEPDVTVITSMITDPSHTTAETAKRLRLARGSPSPSLASSASQLSAPSLKACTTHHD 2220
 Db 2161 PCEPEPDVTVITSMITDPSHTTAETAKRLRLARGSPSPSLASSASQLSAPSLKACTTHHD 2220
 QY 2221 SPDADLIENALLWROEMGNTTRVSEKNVITLDSPEPLHAEGDBERTSVAAEILKSRK 2280
 Db 2221 SPDADLIENALLWROEMGNTTRVSEKNVITLDSPEPLHAEGDBERTSVAAEILKSRK 2280
 QY 2281 PPSALPIWARPDYNPPLLESWKDPYVPPVHGCPLPPTKAPPIPPRRKRTVVLTSNV 2340
 Db 2281 PPSALPIWARPDYNPPLLESWKDPYVPPVHGCPLPPTKAPPIPPRRKRTVVLTSNV 2340

QY 2341 SSALAEALATKTFSSGSSAVDSGTATALPDLASDDGDKGSDVESYSSMPLEGECDPDL 2400
 Db 2341 SSALAEALATKTFSSGSSAVDSGTATGPPDQASDDGDKGSDVESYSSMPLEGECDPDL 2400
 QY 2401 SGDSWSTVSEBASEDVCVCSMSYTWGALITPCAAEESKLPINPLSNSILRHNNVYAIT 2460
 Db 2401 SGDSWSTVSEBASEDVCVCSMSYTWGALITPCAAEESKLPINPLSNSILRHNNVYAIT 2460
 QY 2461 SRSASLRQKKVTFDLQVLDHVRDLVKEMKAKASTVKAKLLSIEACKLTPPHSAKSKF 2520
 Db 2461 SRSASLRQKKVTFDLQVLDHVRDLVKEMKAKASTVKAKLLSIEACKLTPPHSAKSKF 2520
 QY 2521 GYKADVRNLSSRAVNHRSVWEDLLEDETPIDITIMAKSEVFCVQEGKGRKARLIV 2580
 Db 2521 GYKADVRNLSSRAVNHRSVWEDLLEDETPIDITIMAKSEVFCVQEGKGRKARLIV 2580
 QY 2581 FPDLGVRVCEKMAIYDVVSTLPOAVMGSSYGFQSPKORVEFLVNTWKSCKCPMGFSYDT 2640
 Db 2581 FPDLGVRVCEKMAIYDVVSTLPOAVMGSSYGFQSPKORVEFLVNTWKSCKCPMGFSYDT 2640
 QY 2641 RCFDSTVTESDIRVBESIYQCCDLAPEARQALIRSLTERLYIGGPLNLSKGQCGYRRA 2700
 Db 2641 RCFDSTVTESDIRVBESIYQCCDLAPEARQALIRSLTERLYIGGPLNLSKGQCGYRRA 2700
 QY 2701 SGVLTSOQNTLTCYLKATAACRAAKLQDCTWLVNGDDLVICESAGTOEDAAALRAFTE 2760
 Db 2701 SGVLTSOQNTLTCYLKATAACRAAKLQDCTWLVNGDDLVICESAGTOEDAAALRAFTE 2760
 QY 2761 AMTRYAPAPQDPQPPBYDLELITSCSSNVSAHDASGKRVVYLTRDPTTFLARAAMETAR 2820
 Db 2761 AMTRYAPAPQDPQPPBYDLELITSCSSNVSAHDASGKRVVYLTRDPTTFLARAAMETAR 2820
 QY 2821 HTPINSWLGNIIMYAPTLWARMLMTHFPIIAOQLEKALDCQIYGACYSIEPLDLPO 2880
 Db 2821 HTPINSWLGNIIMYAPTLWARMLMTHFPIIAOQLEKALDCQIYGACYSIEPLDLPO 2880
 QY 2881 IBERLHGLAFTHSYPGEINRVSCLKLGVPPLRTWHRARSVRAKLLSQGGAATC 2940
 Db 2881 IBERLHGLAFTHSYPGEINRVSCLKLGVPPLRTWHRARSVRAKLLSQGGAATC 2940
 QY 2941 GRYLFNWAYRTKLTPIPAASOLDISGMFVAGYSGGDIYHLSLRARPRWFLCLLLSV 3000
 Db 2941 GRYLFNWAYRTKLTPIPAASOLDISGMFVAGYSGGDIYHLSLRARPRWFLCLLLSV 3000
 QY 3001 GVGIYLLPNR 3010
 Db 3001 GVGIYLLPNR 3010

RESULT 6
 AAR68622
 ID AAR68622 standard; protein; 3010 AA.
 XX
 AC AAR68622;
 XX
 DT 16-OCT-2003 (revised)
 DT 16-OCT-1995 (first entry)
 XX
 XX HCV protein cleavable with new serine proteinase.
 XX
 XX proteinase; serine; cleavage; hepatitis C virus; HCV.
 XX
 XX Hepatitis C virus; Virus.
 OS
 PH key Location/Qualifiers
 FT Cleavage-site 2419..2420
 FT /note= "Serine protease cleavage site"
 XX
 XX JP06315377-A.
 XX
 XX 15-NOV-1994.
 XX
 XX 06-MAY-1993; 93JP-00105666.

XX PR 06-MAY-1993; 93JP-00105666.
 XX PA (KAEN/) KAENNO K.
 XX PA (SUMQ) SUMITOMO METAL IND LTD.
 XX PA (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
 XX DR WPI; 1995-032330/05.
 XX DR N-PSDB; AAQ80498.
 XX PT New HCV-originated proteinase active substance - used for site-specific
 PT cleavage by an intermolecular reaction and the purification thereof.
 XX PS Disclosure; Page 10-19; 23pp; Japanese.
 XX CC This protein from HCV (hepatitis C virus) (encoded by AAQ80498) is
 CC cleaved between amino acids 2419 and 2420, by a new serine protease,
 CC contg. the sequence of AA68621. The proteinase is purified as a fused
 CC product with the dihydrofolate reductase protein by using a methotrexate
 CC column. It can be used for the development of an inhibitor for HCV
 CC proteinase. (Updated on 16-OCT-2003 to standardise OS field)
 XX SQ Sequence 3010 AA;
 Query Match 95.8%; Score 15332; DB 2; Length 3010;
 Best Local Similarity 95.0%; Pred. No. 0;
 Matches 2860; Conservative 65; Mismatches 85; Indels 0; Gaps 0;
 QY 1 MSTNPKQKTKRNINRRPQVPPGGQIVGGVYLLPRRGPRLGVRATKASRSQPRG 60
 DB 1 MSTNPKQKTKRNINRRPQVPPGGQIVGGVYLLPRRGPRLGVRATKASRSQPRG 60
 QY 61 RQPIPKARRPGRMAOPGVWPLVYGNELGWAGWLLSPRGSPPSWGPTDPRRSRLG 120
 DB 61 RQPIPKARRPGRMAOPGVWPLVYGNELGWAGWLLSPRGSPPSWGPTDPRRSRLG 120
 QY 121 KVIDTLTCGFDLADMGYIFLVGAPLGGAAARALAHGVRLVEDGVNYATGNLPGCSFIFLLA 180
 DB 121 KVIDTLTCGFDLADMGYIFLVGAPLGGAAARALAHGVRLVEDGVNYATGNLPGCSFIFLLA 180
 QY 181 LLSCLTIPASAYEVNVSIGIYHVINDCNSISIVIEAADVIMETPCQVPCVQEGNSRCWV 240
 DB 181 LLSCLTIPASAYEVNVSIGIYHVINDCNSISIVIEAADVIMETPCQVPCVQEGNSRCWV 240
 QY 241 ALTPTLAARNASVPTTIRRHVDLVAGTAAPFCSAMVVDLCSIFLVSQLTPSPRRHET 300
 DB 241 ALTPTLAARNASVPTTIRRHVDLVAGTAAPFCSAMVVDLCSIFLVSQLTPSPRRHET 300
 QY 301 VQDCNCSIPGHVSGRHMAWDMNWSPTTALVVSQLLRIPOAVVDMVAGAHWGLAGLA 360
 DB 301 VQDCNCSIPGHVSGRHMAWDMNWSPTTALVVSQLLRIPOAVVDMVAGAHWGLAGLA 360
 QY 361 YYSWGVNKAULIYALLFAGVDGETHTRGVAGHTISGTSLSFSSGASQKQLVNTNGSW 420
 DB 361 YYSWGVNKAULIYALLFAGVDGETHTRGVAGHTISGTSLSFSSGASQKQLVNTNGSW 420
 QY 421 HINRTALNCNDSLOTGFFAALFYAHKENSNGCPERMASCRPTDFAQGWGPIITYKPNSS 480
 DB 421 HINRTALNCNDSLOTGFFAALFYAHKENSNGCPERMASCRPTDFAQGWGPIITYKPNSS 480
 QY 481 DQRPYCHYAPRCGVFPASQVCGVYCTPSPVGVVGTDRSGVPTYSNGENETDVLIN 540
 DB 481 DQRPYCHYAPRCGVFPASQVCGVYCTPSPVGVVGTDRSGVPTYSNGENETDVLIN 540
 QY 541 NTRPPQGNWFGCTWNNSTGFTKCGPPCNIIGVGNRTLICDTCPRKHPEATYKCGSG 600
 DB 541 NTRPPQGNWFGCTWNNSTGFTKCGPPCNIIGVGNRTLICDTCPRKHPEATYKCGSG 600
 QY 601 PMLTPRCLVDYVYRLWHYVPCTLNFSIFPKVMYVGVVGVVGHRLNAACNWRGCRNLEDNRDS 660
 DB 601 PMLTPRCLVDYVYRLWHYVPCTLNFSIFPKVMYVGVVGVVGHRLNAACNWRGCRNLEDNRDS 660
 QY 661 ELSPLLSTTEWQILPCSFITPALSTGLIHLHQNIVDVQVLYGVGSFAFVSKWEYIL 720

661 ELSPLLSTTEWQILPCSFITPALSTGLIHLHQNIVDVQVLYGVGSFAFVSKWEYIL 720
 QY 721 LLFLLLLADARVCACLWMMLLIACAAALENLVLAASVAGAGHIIISFLVFFCAAAYIKG 780
 DB 721 LLFLLLLADARVCACLWMMLLIACAAALENLVLAASVAGAGHIIISFLVFFCAAAYIKG 780
 QY 781 RLAPGAAYAPYGVWPLLLLLALLPRAVALDEMMAASCGGAVLVGLVFTLSPYKVFLL 840
 DB 781 RLAPGAAYAPYGVWPLLLLLALLPRAVALDEMMAASCGGAVLVGLVFTLSPYKVFLL 840
 QY 841 RLIIWVLYQVITBARAHMOVVPPVNVGRGGRDAIILLTCAVHPELIFDITKLLAILGLPLM 900
 DB 841 RLIIWVLYQVITBARAHMOVVPPVNVGRGGRDAIILLTCAVHPELIFDITKLLAILGLPLM 900
 QY 901 VLOAGITRVPYFVRAOCLIRACMLNKKVAGGHVQVFMKLGALITGYVYVNHULTPLRDA 960
 DB 901 VLOAGITRVPYFVRAOCLIRACMLNKKVAGGHVQVFMKLGALITGYVYVNHULTPLRDA 960
 QY 961 HAGLRDLAVALVEPVFSAMETKVITWGAATAACGDIILGLPVSARRGKHIFLGPADSLEG 1020
 DB 961 HAGLRDLAVALVEPVFSAMETKVITWGAATAACGDIILGLPVSARRGKHIFLGPADSLEG 1020
 QY 1021 QGWELLAPITAYSOOTRGVLGCIITSITGSDKNQVGEVOVSTATQSFATCINGVCWT 1080
 DB 1021 QGWELLAPITAYSOOTRGVLGCIITSITGSDKNQVGEVOVSTATQSFATCINGVCWT 1080
 QY 1081 VYHAGSKTLAGPKPITQMYTNVDLDLVGWQAPPGARSMTPCSCGSSDLYLVTRHADVI 1140
 DB 1081 VYHAGSKTLAGPKPITQMYTNVDLDLVGWQAPPGARSMTPCSCGSSDLYLVTRHADVI 1140
 QY 1141 FVRRGDSRGLSPRPVSYKLGSSGGLLPCPGHVVGVPFAAIVCTRGVAKAVDFIPVES 1200
 DB 1141 FVRRGDSRGLSPRPVSYKLGSSGGLLPCPGHVVGVPFAAIVCTRGVAKAVDFIPVES 1200
 QY 1201 MEITMRSPVTDNSTPFAVPQTQVVAHLHAPTQSGKSTKVPAAVAAQYKVLVNPVAA 1260
 DB 1201 MEITMRSPVTDNSTPFAVPQTQVVAHLHAPTQSGKSTKVPAAVAAQYKVLVNPVAA 1260
 QY 1261 TLGFGAYMSKAHGDIPNIRTVGRTITTTGGSTYTYKFLADGGCSGAYDIILICDECHS 1320
 DB 1261 TLGFGAYMSKAHGDIPNIRTVGRTITTTGGSTYTYKFLADGGCSGAYDIILICDECHS 1320
 QY 1321 TDSTTILIGITVLDQAETAGARLVLTATAPPGSVTVPHNIBBIIIGLNNGEIPFYKAI 1380
 DB 1321 TDSTTILIGITVLDQAETAGARLVLTATAPPGSVTVPHNIBBIIIGLNNGEIPFYKAI 1380
 QY 1381 PIEAIKGRHLIFCHSKKKODELAALKUTGLGNNAVAYRGLDVSVIPIIGDVVVVATDAL 1440
 DB 1381 PIEAIKGRHLIFCHSKKKODELAALKUTGLGNNAVAYRGLDVSVIPIIGDVVVVATDAL 1440
 QY 1441 MTGFTGDPDSVIDNCTVQTVDFSLDPTFTIETTTVPQDAVSRQRRGTRGRSGIYR 1500
 DB 1441 MTGFTGDPDSVIDNCTVQTVDFSLDPTFTIETTTVPQDAVSRQRRGTRGRSGIYR 1500
 QY 1501 FVTPCGRPSGMFDSVILCECVDACAWVELTPASTSVRLRAYLNTPLGIPCODHLEPWS 1560
 DB 1501 FVTPCGRPSGMFDSVILCECVDACAWVELTPASTSVRLRAYLNTPLGIPCODHLEPWS 1560
 QY 1561 VFTGLTHIDAHFLSQTKOAGDNFVLVAYQATVCARAQPPPSQDWKWLIRLKPITHG 1620
 DB 1561 VFTGLTHIDAHFLSQTKOAGDNFVLVAYQATVCARAQPPPSQDWKWLIRLKPITHG 1620
 QY 1621 PTPLLYRLGAVQNEVILTHPTTKYIMACMSADLEVTSTWLVGVVLAALAAAYCLTTGSV 1680
 DB 1621 PTPLLYRLGAVQNEVILTHPTTKYIMACMSADLEVTSTWLVGVVLAALAAAYCLTTGSV 1680
 QY 1681 VIVGRILISGKPAVVPDREVLVYQFDEMEHCASQIPYEQOQMLABQFKQKALGLQAT 1740
 DB 1681 VIVGRILISGKPAVVPDREVLVYQFDEMEHCASQIPYEQOQMLABQFKQKALGLQAT 1740
 QY 1741 KQAEAAAPVFSKWRPALETFWAKHWNFIISGIVYLAGLSTLPGNPATASLMAFASITSP 1800

1741 KQAEAAAPVSEKWRALAEVFNWAKHWNFISSGIQYLGLSTLPGNPALASLMAFTASITSP 1600
1801 LTTQNTLNLGGWVAAQLAPPASAAPVAGIAGAAVGSIGLQKVLVDIILAGYGAGVA 1860
1801 LTTQNTLNLGGWVAAQLAPPASAAPVAGIAGAAVGSIGLQKVLVDIILAGYGAGVA 1860
1861 GALVAFKMGSEVSTEDLAVNLIPALISPGALVGVVCAIILRRHVGPGEAGAVQWNRLLI 1920
1861 GALVAFKMGSEVSTEDLAVNLIPALISPGALVGVVCAIILRRHVGPGEAGAVQWNRLLI 1920
1921 APASRGNHVSPTHYVSPESDAARVTQILSSLTITOLLKELHQMINEDECSGPCGSLRDV 1980
1921 APASRGNHVSPTHYVSPESDAARVTQILSSLTITOLLKELHQMINEDECSGPCGSLRDV 1980
1981 NWICVTLDFKTWLOSKLLPLPGVPLFLSCQBGYKGVWRGDMQITCPCGAQIAGHVK 2040
1981 NWICVTLDFKTWLOSKLLPLPGVPLFLSCQBGYKGVWRGDMQITCPCGAQIAGHVK 2040
2041 NGSMRIVGRTCSNTWHTGTFPINAYTTGCTPSPAPNYGRALWRVAABEYVEVTRVGDPH 2100
2041 NGSMRIVGRTCSNTWHTGTFPINAYTTGCTPSPAPNYGRALWRVAABEYVEVTRVGDPH 2100
2101 VYTGMTTNDVKPCQVPAPPEFFTEVDGVLHRYAPACKPELLREDVTFQGLNQYLWGSOL 2160
2101 VYTGMTTNDVKPCQVPAPPEFFTEVDGVLHRYAPACKPELLREDVTFQGLNQYLWGSOL 2160
2161 PCEPEPDVTLTSMITDPSHITAEAKRLARGSPPLASSASQI-SAPSLKATCTTHD 2220
2161 PCEPEPDVTLTSMITDPSHITAEAKRLARGSPPLASSASQI-SAPSLKATCTTHD 2220
2221 SPDADLIEANLLWRQMGNGNITRESENKVVILDSFEPLHAEGDEREISVAABILKRSK 2280
2221 SPDADLIEANLLWRQMGNGNITRESENKVVILDSFPPIRAVEDERZISVPAEILAKPK 2280
2281 FPSALPIWAPDPYNNPILLESKWDPPYVPPVHGCPLPPTKAPDIPPRKRTWLTESNV 2340
2281 FPSALPIWAPDPYNNPILLESKWDPPYVPPVHGCPLPPTKAPDIPPRKRTWLTESNV 2340
2341 SSALAEIATTFGSSGSAVDSGTATLPLDASDGDGSDVRSYSMPLEGEPPDL 2400
2341 SSALAEIATTFGSSGSAVDSGTATLPLDASDGDGSDVRSYSMPLEGEPPDL 2400
2401 SDGSWSVSEAGEDVVCCSMYSYTWGALITPCAAEESKLPINPLNSLRLHNNMVYATT 2460
2401 SDGSWSVSEAGEDVVCCSMYSYTWGALITPCAAEESKLPINPLNSLRLHNNMVYATT 2460
2461 SRGASLFOKKVTFDRLQVLDHHRDVLKEMKAKASTVAKLLSIEEACKLTPPHSAKSF 2520
2461 SRGASLFOKKVTFDRLQVLDHHRDVLKEMKAKASTVAKLLSIEEACKLTPPHSAKSF 2520
2521 GYGAKDVRLSSRAVNHRSWEDLLEDTETPIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
2521 GYGAKDVRLSSRAVNHRSWEDLLEDTETPIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
2581 FPDGLGVRCERKALYDVVSTLPQAVMGSGYGFQSPQORVEFLVNTWKSCKCPMGFSYDT 2640
2581 FPDGLGVRCERKALYDVVSTLPQAVMGSGYGFQSPQORVEFLVNTWKSCKCPMGFSYDT 2640
2641 RCFDSTVTESDIRVEESIQCCDLAPEARQAIRSLTERLYIGGLTNSKGNCCYRRCRA 2700
2641 RCFDSTVTESDIRVEESIQCCDLAPEARQAIRSLTERLYIGGLTNSKGNCCYRRCRA 2700
2701 SGVLTTSNGNLTCLVKATAACRAKLODCTMLVNGDDLWVCFESAGTQEDAAALRAFTE 2760
2701 SGVLTTSNGNLTCLVKATAACRAKLODCTMLVNGDDLWVCFESAGTQEDAAALRAFTE 2760
2761 AMTRYSAFPGDPPEYDLELITSCSNVSVVAHDASGKRVYVLTDRPTPLARAAMETAR 2820
2761 AMTRYSAFPGDPPEYDLELITSCSNVSVVAHDASGKRVYVLTDRPTPLARAAMETAR 2820
2821 HTPINSWLNIIIMYAPTIAWARMILMTHFFSILLAGEOLEKALDCQIYGACYSIEPLDLPO 2880
2821 HTPINSWLNIIIMYAPTIAWARMILMTHFFSILLAGEOLEKALDCQIYGACYSIEPLDLPO 2880

2881 IIERLHGLSAPTLKSYSPGEINRVASCLIKGLGVPLRTWRHARSVRKALLSQGGRATC 2940
2881 IIERLHGLSAPTLKSYSPGEINRVASCLIKGLGVPLRTWRHARSVRKALLSQGGRATC 2940
2941 GRYLFNNAVRKLKLTIPAPASQDLDSGWFFVAGVSGGDIYHSLSRAPRPFPLCLLLLSV 3000
2941 GRYLFNNAVRKLKLTIPAPASQDLDSGWFFVAGVSGGDIYHSLSRAPRPFPLCLLLLSV 3000
3001 GVGIYLLPNR 3010
3001 GVGIYLLPNR 3010
RESULT 7
ID ABG32451 standard; protein; 3010 AA.
XX AC ABG32451;
XX DT 15-NOV-2002 (first entry)
XX Hepatitis C virus Con 1 isolate polyprotein.
XX HCV; Con 1; adaptive mutation; liver failure; cirrhosis;
XX hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
XX internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
XX Hepatitis C virus.
XX WO200259321-A2.
XX 01-AUG-2002.
XX 16-JUN-2002; 2002WO-EP000526.
XX 23-JAN-2001; 2001US-0263479P.
XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX De Francesco R, Migliaccio G, Paonessa G;
XX WPI; 2002-599793/64.
XX N-PSDB; ABK91411.
XX New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
XX NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
XX ribosome entry site (IRES) region, useful in studying HCV replication and
XX expression.
XX Claim 1; Page 34-36; 69pp; English.
XX The invention relates to nucleic acid molecules comprising altered HCV
XX NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
XX internal ribosome entry site (IRES) region coding for one or more NS3,
XX NS5A, or EMCV IRES mutations, respectively. The location of the mutations
XX are detailed in the specification. Also included are (1) an expression
XX vector comprising a nucleotide sequence coding for the altered nucleic
XX acids, which is transcriptionally coupled to an exogenous promoter; (2) a
XX recombinant cell human hepatoma cell comprising the altered nucleic acids
XX; (3) a recombinant cell produced by introducing into a human hepatoma
XX cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
XX replicon enhanced cell or which containing a functional HCV replicon; (5)
XX an HCV replicon enhanced cells made in the method; and (6) measuring the
XX ability of a compound to affect HCV activity. The HCV replicons and HCV
XX replicon enhanced cells are useful in studying HCV replication and
XX expression, and HCV and host cell interactions, producing HCV RNA and
XX proteins, and providing a system for measuring the ability of a compound
XX to modulate one or more HCV activities e.g. to discover drugs which may
XX treat HCV mediated diseases such as liver failure, cirrhosis and
XX hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
XX polyprotein (comprising the Core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A
XX and NS5B proteins) used as a basis for the adaptive mutations of the

CC	invention	95.7%;	Score 15323;	DB 5;	Length 3010;
XX	Sequence 3010 AA;	Best Local Similarity 94.8%;	Pred. No. 0;	Mismatches 82;	Indels 0;
5Q	Matches 2852;	Conservative 76;	Mismatches 82;	Indels 0;	Gaps 0;
Qy	1	MSNPKPQKTKENTNR	PDVKFPGGGQIVGGVILLPRRGPRLGVRATRKASERSQPRG	60	
Db	1	MSNPKPQKTKENTNR	PDVKFPGGGQIVGGVILLPRRGPRLGVRATRKTSERSQPRG	60	
Qy	61	RRQPIPKARPEGR	QAWAFCGYPMFLYNEGLGWAGWLLSPRGSRRPSWGFTDPRRSRLG	120	
Db	61	RRQPIPKARPEGR	QAWAFCGYPMFLYNEGLGWAGWLLSPRGSRRPSWGFTDPRRSRLG	120	
Qy	121	KVIDTLTCGFADLMGVI	PLVGAFLGGAARALAHGVRLFDGVNYATGNLPGCSRSFIFLLA	180	
Db	121	KVIDTLTCGFADLMGVI	PLVGAFLGGAARALAHGVRLFDGVNYATGNLPGCSRSFIFLLA	180	
Qy	181	LLSLCLATIPAS	AYEVRNVSGIYHYVNDCSNLSIYVEAADVIMETPGCVPCVQSGNSRRQWV	240	
Db	181	LLSLCLATIPAS	AYEVRNVSGIYHYVNDCSNLSIYVEAADVIMETPGCVPCVQSGNSRRQWV	240	
Qy	241	ALPTLAARNAS	VPFTTIRRHVDLLVGTAAFCSAMVYGLCGSIFLVLSOLFTEPSRRHET	300	
Db	241	ALPTLAARNAS	VPFTTIRRHVDLLVGTAAFCSAMVYGLCGSVFLVAQLFTFSPRRHET	300	
Qy	301	VQDNCNSIYFGHVS	GHRYANDMMNWSPTTALIVVLSQLIRIPOAVDMVAGAHGVLAGLA	360	
Db	301	VQDNCNSIYFGHVS	GHRYANDMMNWSPTTALIVVLSQLIRIPOAVDMVAGAHGVLAGLA	360	
Qy	361	YYSVGVGNWAKVLI	VAILEPAGVDGSEHTTGRVAGHTTSGFTSLPSSGASOKIQLVNTNGSW	420	
Db	361	YYSVGVGNWAKVLI	VAILEPAGVDGSEHTTGRVAGHTTSGFTSLPSSGASOKIQLVNTNGSW	420	
Qy	421	HINRTALMND	SLQGTFFALFYAHKNSSGCPERMASCRPIQWPAQGWGPIYTKPNSG	480	
Db	421	HINRTALMND	SLQGTFFALFYAHKNSSGCPERMASCRPIQWPAQGWGPIYTKPNSG	480	
Qy	481	DQPYCHYAPR	PCGVVPASQVCPVCFTPSPVVVGTTRDSRGVTVSGENETDVMLLN	540	
Db	481	DQPYCHYAPR	PCGVVPASQVCPVCFTPSPVVVGTTRDSRGVTVSGENETDVMLLN	540	
Qy	541	NTRPQGNWEGCT	WNNSGTGFKTCGGPPCNIGGVGNRTLICPTDCPRKPEATYTKCGSG	600	
Db	541	NTRPQGNWEGCT	WNNSGTGFKTCGGPPCNIGGVGNRTLICPTDCPRKPEATYTKCGSG	600	
Qy	601	PWLTPRC	LVDYPIRLWHYPCVTLNFSIFKVRMYGVGVHRLNAACNWTGRGCMLEDRS	660	
Db	601	PWLTPRC	LVDYPIRLWHYPCVTLNFSIFKVRMYGVGVHRLNAACNWTGRGCMLEDRS	660	
Qy	661	ELSPILLSTTE	WQILPCAFITLPAISLCLIHHLQNIYDVQVLYGVGSAPVSEALKWEYIL	720	
Db	661	ELSPILLSTTE	WQILPCAFITLPAISLCLIHHLQNIYDVQVLYGVGSAPVSEALKWEYIL	720	
Qy	721	LIPLLLADAR	VACALWMLLIAQAEAALENVLVNAASVAGAHGILSLVFFCAAWYIKG	780	
Db	721	LIPLLLADAR	VACALWMLLIAQAEAALENVLVNAASVAGAHGILSLVFFCAAWYIKG	780	
Qy	781	RLAPGAAYAFY	GVWPFLLILLALPPRAVALDREMAASC GGAVLVGLVFLILSPYKVFILT	840	
Db	781	RLVPGAAYALY	GVWPFLLILLALPPRAYANDREMAASC GGAVFVGLIILLTSPHYKFLA	840	
Qy	841	RLIWMLOYEIT	TAEAHMQVWPEI NVRGRODAIILLTCVAVHPEITFDITKLLAILGLPM	900	
Db	841	RLIWMLOYEIT	TAEAHMQVWPEI NVRGRODAIILLTCVAVHPEITFDITKLLAILGLPM	900	
Qy	901	VLQAGITRPY	FVRAQGLTRACWLVRKVAGGHYQVFWKLCALGTCTVYVNHLPFLRDWA	960	
Db	901	VLQAGITRPY	FVRAQGLTRACWLVRKVAGGHYQVFMKLAALGTCTVYVNHLPFLRDWA	960	
Qy	961	HAGIRDLAV	AVFPVFSAMETKVI TWGADTAAACGDIILGLPVSARRGKEILGPADSLEG	1020	

Dd 2041 NGSRIIVGERTCSNTHGTPINAYTTGCTPSPAPNYSRALWRVAABEYVEVTRVGDFH 2100
 Qy 2101 YVTGMTTNNVKPCQVPAPEPTEVDGVRLEHYAPACKPIAREDVTFQVGLNOVLVSOL 2160
 Dd 2101 YVTGMTTNNVKPCQVPAPEPTEVDGVRLEHYAPACKPIAREDVTFQVGLNOVLVSOL 2160
 Qy 2161 PCEPEPDTVLTSMLTDSHITAEAKRRLARGSPPSLASSASQLSAPSKATCTTHD 2220
 Dd 2161 PCEPEPDTVLTSMLTDSHITAEAKRRLARGSPPSLASSASQLSAPSKATCTTHD 2220
 Qy 2221 SPDADLIENILWRQMGENTRVESENKVLILDSFELHAGDBEREISVAABILRSRK 2280
 Dd 2221 SPDADLIENILWRQMGENTRVESENKVLILDSFELHAGDBEREISVAABILRSRK 2280
 Qy 2281 PPSALPIWARPDYNPPLLESWKDPDYPVVGHCPLPPTKAPPDPPRRKRTVWLTESNV 2340
 Dd 2281 PPSALPIWARPDYNPPLLESWKDPDYPVVGHCPLPPTKAPPDPPRRKRTVWLTESNV 2340
 Qy 2341 SSALAEATKTFGSGSSAVSDGTATLPDASDDGDKGSDVESYSPMPLEGEFGDPL 2400
 Dd 2341 SSALAEATKTFGSGSSAVSDGTATLPDASDDGDKGSDVESYSPMPLEGEFGDPL 2400
 Qy 2401 SDGWSVTSSEASSEDVCCSYTWGTGALITPCAARESKLPINPLNSLIRHNNVYATT 2460
 Dd 2401 SDGWSVTSSEASSEDVCCSYTWGTGALITPCAARESKLPINPLNSLIRHNNVYATT 2460
 Qy 2461 SRSASLRQKVTDFRLQVLDHRYDLVKEMKAKASTVAKLLSTEEACKLTPPHSAKSF 2520
 Dd 2461 SRSASLRQKVTDFRLQVLDHRYDLVKEMKAKASTVAKLLSTEEACKLTPPHSAKSF 2520
 Qy 2521 GYGAKDVNLSRAVNHRSVWDLLEDTETPIDTIMAKSEVFCVQBEKGRKPARLIV 2580
 Dd 2521 GYGAKDVNLSRAVNHRSVWDLLEDTETPIDTIMAKSEVFCVQBEKGRKPARLIV 2580
 Qy 2581 PPDGLVRVCEKXALVDVYSTLPQAVMGSSYGFQSPKORVELVNTWKSCKCPMGFSYDT 2640
 Dd 2581 PPDGLVRVCEKXALVDVYSTLPQAVMGSSYGFQSPKORVELVNTWKSCKCPMGFSYDT 2640
 Qy 2641 RCFDSTVTESDIRVEESIYQCCDLAPPEARQAIRSLTERLYIGGPLTNSKGQNCYVRCRA 2700
 Dd 2641 RCFDSTVTESDIRVEESIYQCCDLAPPEARQAIRSLTERLYIGGPLTNSKGQNCYVRCRA 2700
 Qy 2701 SGVLTSCGNLTLYKATAACRAAKLQDCNVLVNGDDLVIICRSAGTQEDAAIRAPTE 2760
 Dd 2701 SGVLTSCGNLTLYKATAACRAAKLQDCNVLVNGDDLVIICRSAGTQEDAAIRAPTE 2760
 Qy 2761 AMTRYSAPPDGPPEYDLELITSCSNVVAHDASGRVYVLTDRDPTPLARAWEETAR 2820
 Dd 2761 AMTRYSAPPDGPPEYDLELITSCSNVVAHDASGRVYVLTDRDPTPLARAWEETAR 2820
 Qy 2821 HTPNSWIGNIMYAPTJWARMILMTHFFSILLAOEQLKALDQIYGACYSIBPLDLPQ 2880
 Dd 2821 HTPNSWIGNIMYAPTJWARMILMTHFFSILLAOEQLKALDQIYGACYSIBPLDLPQ 2880
 Qy 2881 IIERHLGLSATLHSYSGEINRVASCLKLGVPPLRTWRHRSVRAKLISQGGRAATC 2940
 Dd 2881 IIERHLGLSATLHSYSGEINRVASCLKLGVPPLRTWRHRSVRAKLISQGGRAATC 2940
 Qy 2941 GRYLFNNAVTRTKLTIPIPAASQDLSCWFFVAGYSGGDIYHSLGRARPRWFFPLCILLISV 3000
 Dd 2941 GRYLFNNAVTRTKLTIPIPAASQDLSCWFFVAGYSGGDIYHSLGRARPRWFFPLCILLISV 3000
 Qy 3001 GVGIVLLNLR 3010
 Dd 3001 GVGIVLLNLR 3010

RESULT 8

ABG32460

ID ABG32460 standard; protein; 3010 AA.

XX

AC ABG32460;

XX

DT 15-NOV-2002 (first entry)
 XX Hepatitis C virus Con 1 isolate polyprotein mutant #9.
 DE HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; muten;
 KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
 KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
 XX Hepatitis C virus.
 OS Synthetic.
 XX Key Location/Qualifiers
 FH Misc-difference 2199 /note= "Wild-type Ala substituted by Thr"
 FT
 FT WO200259321-A2.
 XX 01-AUG-2002.
 XX 16-JAN-2002; 2002WO-EP000526.
 XX 23-JAN-2001; 2001US-0263479P.
 XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
 XX De Francesco R, Migliaccio G, Paonessa G;
 XX WPI; 2002-599793/64.
 XX New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
 PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
 PT ribosome entry site (IRES) region, useful in studying HCV replication and
 PT expression.
 XX Claim 1; Page; 69pp; English.
 CC The invention relates to nucleic acid molecules comprising altered HCV
 CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
 CC internal ribosome entry site (IRES) region coding for one or more NS3,
 CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
 CC are detailed in the specification. Also included are (1) an expression
 CC vector comprising a nucleotide sequence coding for the altered nucleic
 CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
 CC recombinant cell human hepatoma cell comprising the altered nucleic acids
 CC ; (3) a recombinant cell produced by introducing into a human hepatoma
 CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
 CC replicon enhanced cell or which containing a functional HCV replicon; (5)
 CC an HCV replicon enhanced cells made in the method; and (6) measuring the
 CC ability of a compound to affect HCV activity. The HCV replicons and HCV
 CC replicon enhanced cells are useful in studying HCV replication and
 CC expression, and HCV and host cell interactions, producing HCV RNA and
 CC proteins, and providing a system for measuring the ability of a compound
 CC to modulate one or more HCV activities e.g. to discover drugs which may
 CC treat HCV mediated diseases such as liver failure, cirrhosis and
 CC hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
 CC polyprotein (comprising the Core, E1, E2, P7, NS3, NS4A, NS4B, NS5A
 CC and NS5B proteins), NS5A mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the HCV sequence appearing as ABG32451 and the information in claim
 CC 1
 XX Sequence 3010 AA;

Query Match 95.7%; Score 15319; DB 5; Length 3010;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 2851; Conservative 76; Mismatches 83; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRTKNTNRPRDVKFGGQIVGGVYLLPRGPRPLGVTRATKASERSQPRG 60

Dd 1 MSTNPKPQRTKNTNRPRDVKFGGQIVGGVYLLPRGPRPLGVTRATKASERSQPRG 60
 61 RROPIKARPEGEFAKACPGYFWPLVNEGLHAGWLLSPGSGSPKGGTDPDRRNLG 120

Db 61 RQPIKQPBGRGAWAQGYFWPIYXNEGGLGWAGWLLSPRGSRPSWGTDTPRRSRNLG 120
Qy 121 KVIDTLTCGFADLMGYIPLVGAPLGAARALAHGVRLVEDGVNYATGNLPGCSFSLFLIA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGAARALAHGVRLVEDGVNYATGNLPGCSFSLFLIA 180
Qy 181 LLSCLTIPASAYEVNRVSGIYHVTNDCSNSSIVYEAADVIMETPGCPVCQVCGNSRRCWV 240
Db 181 LLSCLTIPASAYEVNRVSGIYHVTNDCSNSSIVYEAADMIMETPGCPVCVRENNSRCWV 240
Qy 241 ALPTFLAARNASVPTTIRRHVDLLVGTAAFCASMYVGLDCSIFLVSOLFPTSPRRHET 300
Db 241 ALPTFLAARNASVPTTIRRHVDLLVGTAAFCASMYVGLDCSIFLVSOLFPTSPRRHET 300
Qy 301 VODCNCSIPYGHVSGRMAWDMWMSPTTALVVSQILRIPOAVDMVAGAHGVLGIA 360
Db 301 VODCNCSIPYGHVTGRMAWDMWMSPTTALVVSQILRIPOAVDMVAGAHGVLGIA 360
Qy 361 YYSMVGNMAKVLIVALLFAGVDGETHTTGRVAGHTTSGFTSLFSSGASOKIQLVNTNGSW 420
Db 361 YYSMVGNMAKVLIVMLLPAGVDGTVTGTMAKNTLGTSLFSPGSSQKIQLVNTNGSW 420
Qy 421 HINRTALCNDLSIOTGFPAALFYAHKFNSSGPERMASCRPIDWFAQGWGPITYTKPNSS 480
Db 421 HINRTALCNDLSN"GFALAALFYVHKFNSSGPERMASCSPIDAFAQGWGPITYNESHSS 480
Qy 481 DORPYCWHYAPRPGVVPASOVCGPVYCTPSPVYVGTDRSGVPTYSWGENETDMLIN 540
Db 481 DORPYCWHYAPRPGVVPASOVCGPVYCTPSPVYVGTDRSGVPTYSWGENETDMLIN 540
Qy 541 NTRPQGNWFGCTWMNSTGFTKTCGPPPCNIGVGNRTLCPTDPRKHPEATYTKCGSG 600
Db 541 NTRPQGNWFGCTWMNSTGFTKTCGPPPCNIGVGNRTLCPTDPRKHPEATYTKCGSG 600
Qy 601 PMLTRCLVDYRLLWHYPCNTLNFSLFKVMYVGVGHEHLNACNWTGRONCLDRDRS 660
Db 601 PMLTRCLVDYRLLWHYPCNTLNFSLFKVMYVGVGHEHLNACNWTGRONCLDRDRS 660
Qy 661 ELSPLLLSTTEQWILPCAPTLPALSTGLIHLHONIVDVQYLYGVCSAPVSFAIKWEXIL 720
Db 661 ELSPLLLSTTEQWILPCAPTLPALSTGLIHLHONIVDVQYLYGVCSAPVSFAIKWEXIL 720
Qy 721 LFLLLADARVACALWMLLIIAQBAALLENLVVJNAASVAGAHGILSLFLVFFCAAWYIK 780
Db 721 LFLLLADARVACALWMLLIIAQBAALLENLVVJNAASVAGAHGILSLFLVFFCAAWYIK 780
Qy 781 RLAPGAAYAFYGVWPLIALLLPBRAYALDEMAASCAGVLVGLVLTSPYKVKPLT 840
Db 781 RLVPGAAYALYGVWPLIALLLPBRAYAMDEMAASCAGVAVFVGLIILTLSPHYKFLFLA 840
Qy 841 RLIMWLQYFIRAFAHMQWVPLNVGRGDDAIILLTCAVHPPELLFDITKLLAILGLPM 900
Db 841 RLIMWLQYFIRAFAHMQWVPLNVGRGDDAVILLTCAIHPPELLFTITKILLAILGLPM 900
Qy 901 VLQAGITRVPYVRAAGLIRACMLVRKVAGHYVQWFMKIGALGTYYVYNHLTPLRWA 960
Db 901 VLQAGITRVPYVRAAGLIRACMLVRKVAGHYVQWFMKLAALCTYVYVHLTPLRWA 960
Qy 961 HAGLRDLAVAPVPVFSAMETKVIITWGDATACGDIILGLPVSARRGKEIFLGPADSLEG 1020
Db 961 HAGLRDLAVAPVPVFSAMETKVIITWGDATACGDIILGLPVSARRGKEIFLGPADSLEG 1020
Qy 1021 QWRLLAPITAYSOOTRQVILGCIITSLTGRDKNOVEGEVQVYSTATQSFATCINGVCWT 1080
Db 1021 QWRLLAPITAYSOOTRQVILGCIITSLTGRDKNOVEGEVQVYSTATQSFATCINGVCWT 1080
Qy 1081 VTHGAGSKTAGPKPITOMYTNVDLVLWGQAPPCARSMTPCSCSSDLYLVTRHADVI 1140
Db 1081 VTHGAGSKTAGPKPITOMYTNVDQDLVGWQAPPCARSMTCTCCSSDLYLVTRHADVI 1140
Qy 1141 PVRRRGDSRGLLSRPRVSYLKGSGGP"LLCPSGHVGVFPRAVCTRGVAKADVIPVES 1200
Db 1141 PVRRRGDSRGLLSRPRVSYLKGSGGP"LLCPSGHVGFIRAAVCTRGVAKADVIPVES 1200

Qy 1201 METTMRSPVPTDNSPPPAVPOTPOVAHLHAFTGSKSKTKVPAAVAAOQYKVLVNLPSVAA 1260
Db 1201 METTMRSPVPTDNSPPPAVPOTPOVAHLHAFTGSKSKTKVPAAVAAOQYKVLVNLPSVAA 1260
Qy 1261 TLGFGAYMSKAHGDIPNIRTGVRTITTTGSGTSTYSTYKGFADGGCGGAYDIIICDECHS 1320
Db 1261 TLGFGAYMSKAHGDIPNIRTGVRTITTTGAPITYSTYKGFADGGCGGAYDIIICDECHS 1320
Qy 1321 TDSITLIGIYVLOQAEATAGARLVVLAATATPPGSVTPVPHNIEIGLSNNGEIPFYGKAI 1380
Db 1321 TDSITLIGIYVLOQAEATAGARLVVLAATATPPGSVTPVPHNIEEVALSSTGEIPFYGKAI 1380
Qy 1381 PIETAKGRHILIFCHSKKKDELAALKTGLGINAVAYYRGJLDSVIPPIDGVVVVATDAL 1440
Db 1381 PIETAKGRHILIFCHSKKKDELAALKTGLGINAVAYYRGJLDSVIPPIDGVVVVATDAL 1440
Qy 1441 MTGFTGDFSDVIDCNTCVTQVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRSGIYR 1500
Db 1441 MTGFTGDFSDVIDCNTCVTQVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRSGIYR 1500
Qy 1501 FVTPGERPSGMPDSSVLCEDYDAGCANYELIPATSVRLRAYLNTPLPVCODHLEPWES 1560
Db 1501 FVTPGERPSGMPDSSVLCEDYDAGCANYELIPATSVRLRAYLNTPLPVCODHLEPWES 1560
Qy 1561 VPTGLTHIDAHFLSOTKQAGDNFVLYVAYQATVCARAQAPPPSWDMMKCLIRLKPTLHG 1620
Db 1561 VPTGLTHIDAHFLSOTKQAGDNFVLYVAYQATVCARAQAPPPSWDMMKCLIRLKPTLHG 1620
Qy 1621 PTPLLYRLGAVQNEVILTHPTIKYIMACMSADLEVTTSTWLVGVGLAALAAYCLITGTSV 1680
Db 1621 PTPLLYRLGAVQNEVILTHPTIKYIMACMSADLEVTTSTWLVGVGLAALAAYCLITGTSV 1680
Qy 1681 VIVGRILLSGKPAVVPDREVLYQBFDEMEBCASOLPHYEQCKQLAABQPKQKALGLQAT 1740
Db 1681 VIVGRILLSGKPAVVPDREVLYQBFDEMEBCASELPHYEQCKQLAABQPKQKALGLQAT 1740
Qy 1741 KOAEAAAAPVBSKVRALFTFWAKHWNFI SGIOYLAGLSTLPGNPAIASLMAFTASITSP 1800
Db 1741 KOAEAAAAPVBSKVRALFTFWAKHWNFI SGIOYLAGLSTLPGNPAIASLMAFTASITSP 1800
Qy 1801 LTTQNTLLFNILGCVAAQLAPPSSAASAFVAGIAGAAVSGIGLKVLVDILAGYGAGVA 1860
Db 1801 LTTQNTLLFNILGCVAAQLAPPSSAASAFVAGIAGAAVSGIGLKVLVDILAGYGAGVA 1860
Qy 1861 GALVAFKVMASGEVSTEDLVNLLPAILSPGALVVGWCAAILRHRVGPGECAVQWMLRI 1920
Db 1861 GALVAFKVMASGEVSTEDLVNLLPAILSPGALVVGWCAAILRHRVGPGECAVQWMLRI 1920
Qy 1921 AFASRGHVSPTHVVPESDAAARVTQILSSLTITOLLKRLHWINEDCSTPCSGSWLRDV 1980
Db 1921 AFASRGHVSPTHVVPESDAAARVTQILSSLTITOLLKRLHWINEDCSTPCSGSWLRDV 1980
Qy 1981 WDMICTVLTDPKTMLQSKLLPRLPGVPPFLSCORGKYGWVRGDMQTTCPGCAIAGHVK 2040
Db 1981 WDMICTVLTDPKTMLQSKLLPRLPGVPPFLSCORGKYGWVRGDMQTTCPGCAIAGHVK 2040
Qy 2041 NGSMRIIVGPRICSNTHGTPFINAYTTGCTPSPAPNYSRALWVAABEYVEVTRVGDPH 2100
Db 2041 NGSMRIIVGPRICSNTHGTPFINAYTTGCTPSPAPNYSRALWVAABEYVEVTRVGDPH 2100
Qy 2101 YVTGMTTNDVPCQVPAPEFFTEVDGVRLEHRYAPACKFILLREDDVTPOGLNQLVLSQ 2160
Db 2101 YVTGMTTNDVPCQVPAPEFFTEVDGVRLEHRYAPACKFILLREDDVTPOGLNQLVLSQ 2160
Qy 2161 PCEPEPDVVLITSMLTDPHSHTAETAKERLARGSPSSSLASSASQLSAPSLKATCTTHD 2220
Db 2161 PCEPEPDVVLITSMLTDPHSHTAETAKERLARGSPSSSLASSASQLSAPSLKATCTTHD 2220
Qy 2221 SPDADLLEANLLWQEMGNTTRVSEKENVILDSFPEPLHAEGBERISVAAEILRSRK 2280
Db 2221 SPDADLLEANLLWQEMGNTTRVSEKENVILDSFPEPLHAEGBERISVAAEILRSRK 2280

```
QY 2281 FFSALPIWAPDPYNPZLLBSWKDPDYVPPVHVCPLPPTKAPDIPPPRRKRTVILTESNV 2340
Db 2281 FFRAMPDIWAPEDYNPZLLBSWKDPDYVPPVHVCPLPPTKAPDIPPPRRKRTVILTESNV 2340
QY 2341 SSALAELAKTFFGSSSSAVDSGTATAPDLASDDGKSDVESYSMPPLGEPDGPL 2400
Db 2341 SSALAELAKTFFGSSSSAVDSGTATAPDLASDDGKSDVESYSMPPLGEPDGPL 2400
QY 2401 SDGSWTSVEEASEDVVCCSMSTWTGALITPCAAEESKLPINPLNSLLRHHNVYAT 2460
Db 2401 SDGSWTSVEEASEDVVCCSMSTWTGALITPCAAEETKLPINALNSLLRHHNVYAT 2460
QY 2461 SRASLRQKKVTPDRLOVLDHYRVLKEMKAKASTVKAKLLSIEACKLTTPHSAKSF 2520
Db 2461 SRASLRQKKVTPDRLOVLDHYRVLKEMKAKASTVKAKLLSVEACKLTTPHSAKSF 2520
QY 2521 GYGAQDVRLSSPANNHISVWEDLEDTETPIDTITIMAKSEVFCVQPKGGRKPARLIV 2580
Db 2521 GYGAQDVRLSSKAVNHISVWKLLEDTETPIDTITIMAKNEVFCVQPKGGRKPARLIV 2580
QY 2581 FPDGLGVRCERKALXDVVSTLPQAVMGSSYGFOYSPKQRFELVNTWKSCKCPMGFSYDT 2640
Db 2581 FPDGLGVRCERKALXDVVSTLPQAVMGSSYGFOYSPGQRFELVNAWAKKCPMGFAYDT 2640
QY 2641 RCFDSTVTSDDRVRVESIYQCDLAPARQAIRSLTERLYIGPITNSKGQCYRRORA 2700
Db 2641 RCFDSTVTENDIRVESIYQCDLAPARQAIRSLTERLYIGPITNSKGQCYRRORA 2700
QY 2701 SGVLTSCGNTLTCYLKATAACRAAKLQDCTMLVNGDDVVICESAGTOEDAAALRAFTE 2760
Db 2701 SGVLTSCGNTLTCYLKAAAACRAAKLQDCTMLVCGDDLVVICESAGTOEDASLRAFTE 2760
QY 2761 AMTRYSAAPGDPQPEYDLELITSCSSNVSVADHASKRXYVLTDRPTTPLARAAMETAR 2820
Db 2761 AMTRYSAAPGDPKPEYDLELITSCSSNVSVADHASKRXYVLTDRPTTPLARAAMETAR 2820
QY 2821 HPIFNSWLGNIIMYAPTLWARMILMTHPSSILLAQBOLKALDCQIYGACYIEPLDLPO 2880
Db 2821 HPIFNSWLGNIIMYAPTLWARMILMTHPSSILLAQBOLKALDCQIYGACYIEPLDLPO 2880
QY 2881 IIERLHGLSAFTLHYSPOEINRVASCLRLKGVPELRTWHRARSVRKLLSQGGAATC 2940
Db 2881 IIFQLHGLSAFSLHYSPOEINRVASCLRLKGVPLRVWHRARSVRKLLSQGGAATC 2940
QY 2941 GYLYLNWAVTKLKTPTPAASQLDLSGMFVAGYGGDTIYHLSRARPRWFPLCLLLSV 3000
Db 2941 GYLYLNWAVTKLKTPTPAASQLDLSGMFVAGYGGDTIYHLSRARPRWFPLCLLLSV 3000
QY 3001 GVGIVLLPNR 3010
Db 3001 GVGIVLLPNR 3010

RESULT 9
ABG32454
ID ABG32454 standard; protein; 3010 AA.
XX AC ABG32454;
XX XX
DT 15-NOV-2002 (first entry)
DE Hepatitis C virus Con 1 isolate polyprotein mutant #3.
XX XX
KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; muten;
KW Hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
OS Hepatitis C virus.
OS Synthetic.
XX XX
FH Key Location/Qualifiers
FT Misc-difference 1347
FT /note= "Wild-type Ala substituted by Thr"
```

```
XX WO200259321-A2.
PN 01-AUG-2002.
PD 16-JAN-2002; 2002WO-EP000526.
PF 23-JAN-2001; 2001US-0263479P.
XX (RICE-) IST RICECHE BIOL MOLECOLARE ANGELETTI.
PA De Francesco R, Migliaccio G, Paonessa G;
XX WPI; 2002-599793/64.
XX New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT ribosome entry site (IRES) region, useful in studying HCV replication and
PT expression.
XX Claim 1; Page; 69pp; English.
XX The invention relates to nucleic acid molecules comprising altered HCV
CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC internal ribosome entry site (IRES) region coding for one or more NS3,
CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC are detailed in the specification. Also included are (1) an expression
CC vector comprising a nucleotide sequence coding for the altered nucleic
CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC recombinant cell human hepatoma cell comprising the altered nucleic acids
CC; (3) a recombinant cell produced by introducing into a human hepatoma
CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC replicon enhanced cell or which containing a functional HCV replicon; (5)
CC an HCV replicon enhanced cells made in the method; and (6) measuring the
CC ability of a compound to affect HCV activity. The HCV replicons and HCV
CC replicon enhanced cells are useful in studying HCV replication and
CC expression, and HCV and host cell interactions, producing HCV RNA and
CC proteins, and providing a system for measuring the ability of a compound
CC to modulate one or more HCV activities e.g. to discover drugs which may
CC treat HCV mediated diseases such as liver failure, cirrhosis and
CC hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
CC polyprotein (comprising the Core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A
CC and NS5B proteins), NS3 mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the HCV sequence appearing as ABG32451 and the information in claim
CC 1
XX SQ Sequence 3010 AA;
Query Match 95.7%; Score 15319; DB 5; Length 3010;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 2851; Conservative 76; Mismatches 83; Indels 0; Gaps 0;
QY 1 MSTNEPKQKTKRNTNRRPDVXPGGGQIVGGVYLLPRGPRGLGVRAATKASRSQPRG 60
Db 1 MSTNEPKQKTKRNTNRRPDVXPGGGQIVGGVYLLPRGPRGLGVRAATKTSRSQPRG 60
QY 61 RRQIPKARPEGRAWAQPGYVPWPLYGNEGIGWAGWLLSPRGSPPSMGPTDPRRSNIG 120
Db 61 RRQIPKARPEGRAWAQPGYVPWPLYGNEGIGWAGWLLSPRGSPPSMGPTDPRRSNIG 120
QY 121 KVIDTLGCFADLMGYIPLVGAPLGGAAALAHGVRVLEDCGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLGCFADLMGYIPLVGAPLGGAAALAHGVRVLEDCGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASAYEVNRVSGIYHVINDCSNNSIIVYEAADVIMHTEGCVPCVQBGNSRCWV 240
Db 181 LLSCLTTPASAYEVNRVSGIYHVINDCSNNSIIVYEAADVIMHTEGCVPCVQBGNSRCWV 240
QY 241 ALTTFLAARNASVPTTIRRHVDLLVGTAAFCSAMVYVGDLCGSIFLYSQLFTPSRRHET 300
Db 241 ALTTFLAARNASVPTTIRRHVDLLVGTAAFCSAMVYVGDLCGSIFLYSQLFTPSRRHET 300
```

QY 301 VQDCNCSIYPGHVSGHMAWDMNWSPTTALVVSOLLRIPOAVVDMVAGAHGVLAGIA 360
Db 301 VQDCNCSIYPGHVSGHMAWDMNWSPTTALVVSOLLRIPOAVVDMVAGAHGVLAGIA 360
QY 361 YYSMVGNMAKVLIVALLFAGVDEHTTGRVAGHTTSFTSLFSSGASOKIOLVNTNGSW 420
Db 361 YYSMVGNMAKVLIVALLFAGVDEHTTGRVAGHTTSFTSLFSSGASOKIOLVNTNGSW 420
QY 421 HLNRTALNCNDSIQGFFAALFAHKNSSGCPERMASCRPIDWFAQAGRPITYTKPNSS 480
Db 421 HLNRTALNCNDSINTGFLAALFYVHKFNSSGCPERMASCPIDAFQAQAGRPITYNESHSS 480
QY 481 DQAPYCWYHAPRCGVVPASQVGVVCTPSPVVPVVTGTRGVPTYSNGENETDVMLIN 540
Db 481 DQAPYCWYHAPRCGVVPAAQVGVVCTPSPVVPVVTGTRGVPTYSNGENETDVMLIN 540
QY 541 NTRPPQGNWFGCTWMNSTGFTKICGPPCNIIGVGNRTLICTDCFRKHPKPEATYTKCGSG 600
Db 541 NTRPPQGNWFGCTWMNSTGFTKICGPPCNIIGVGNRTLICTDCFRKHPKPEATYTKCGSG 600
QY 601 PWLTPRCVLVDYPRYLWHYECTLNFSTFKVMYVGVGVEHRLNAACNWRGERCNLEDRDRS 660
Db 601 PWLTPRCVLVDYPRYLWHYECTLNFSTFKVMYVGVGVEHRLNAACNWRGERCNLEDRDRS 660
QY 661 ELSPLLSTTEWQIIPCAFTTLPALSTGLIHHQNIIVQXLYGVGSAPVSPFAIKWXYIL 720
Db 661 ELSPLLSTTEWQIIPCAFTTLPALSTGLIHHQNIIVQXLYGVGSAPVSPFAIKWXYIL 720
QY 721 LFLLLADARVACACIWMMLLIAQAALENVLNNAASVAGAHGILSLVFFCAAWYIKG 780
Db 721 LFLLLADARVACACIWMMLLIAQAALENVLNNAASVAGAHGILSLVFFCAAWYIKG 780
QY 781 RIAPGAAYAFYGVWPELLLLALPPRAVALDREMAASCGGAVLVLFTLSPYKVFIL 840
Db 781 RIAPGAAYAFYGVWPELLLLALPPRAVALDREMAASCGGAVLVLFTLSPYKVFIL 840
QY 841 RLIMWLQYFIITAEAHMQVWVPLNVGRGDRAIILLTCAVHPELIDITKLLAILGLFELM 900
Db 841 RLIMWLQYFIITAEAHMQVWVPLNVGRGDRAIILLTCAVHPELIDITKLLAILGLFELM 900
QY 901 VLQAGITVPVYFVRAQGLIRACMLVRKVGAGHYVQVFMKLGALGCTVYVXHLTPLRDWA 960
Db 901 VLQAGITVPVYFVRAQGLIRACMLVRKVGAGHYVQVFMKLGALGCTVYVXHLTPLRDWA 960
QY 961 HAGLRDLAVAVEPVFSDMETKVIITWGDATACGDIILGLPVSARRGKEIFLGPADSLEG 1020
Db 961 HAGLRDLAVAVEPVFSDMETKVIITWGDATACGDIILGLPVSARRGKEIFLGPADSLEG 1020
QY 1021 QGWELLADITAYSQOTRGVLGCIITSLTGRDNQVGEVQVWVSTATQSFATCINGVCWT 1080
Db 1021 QGWELLADITAYSQOTRGVLGCIITSLTGRDNQVGEVQVWVSTATQSFATCINGVCWT 1080
QY 1081 VYHAGSKTLGPKGPIQMTYNTVDLDVAGWQAPPGASMTPCSCGSDILVYTHADVI 1140
Db 1081 VYHAGSKTLGPKGPIQMTYNTVDLDVAGWQAPPGASMTPCSCGSDILVYTHADVI 1140
QY 1141 PVRRRGSGRSLGPRVSYLKGSGGLLCPSGHVGVFRAAVCTRGVAKAVDFIPVES 1200
Db 1141 PVRRRGSGRSLGPRVSYLKGSGGLLCPSGHVGVFRAAVCTRGVAKAVDFIPVES 1200
QY 1201 METTWRSVFTDNSTPRAVQPTQVAHLHATGSGKSTKVPAAVAAQYKVLVNPVAA 1260
Db 1201 METTWRSVFTDNSTPRAVQPTQVAHLHATGSGKSTKVPAAVAAQYKVLVNPVAA 1260
QY 1261 TLGFGAYMSKAGHDIPNRTGVRTITTCGSLITYSTYKGFADGGCGGAYDIIICDRCHS 1320
Db 1261 TLGFGAYMSKAGHDIPNRTGVRTITTCGSLITYSTYKGFADGGCGGAYDIIICDRCHS 1320
QY 1321 TDSITILGIGTVLQDAETAGARLVYLATATPPGSVTVPHPNIEBTGNSNGEIPFYKAI 1380
Db 1321 TDSITILGIGTVLQDAETAGARLVYLATATPPGSVTVPHPNIEBTGNSNGEIPFYKAI 1380
QY 1381 PIEAKGGRHLIFCHSKKKCDELAALKUTGLGLNAVAYYRGLDVSVIPP2GDVVVVATDAL 1440

Db 1381 PIEAKGGRHLIFCHSKKKCDELAALKUTGLGLNAVAYYRGLDVSVIPP2GDVVVVATDAL 1440
QY 1441 MTGFTGDFSDVIDNCTVOTQVDFSLDPTFTTETTVTPQDAVSRQRGRGSGIYR 1500
Db 1441 MTGFTGDFSDVIDNCTVOTQVDFSLDPTFTTETTVTPQDAVSRQRGRGSGIYR 1500
QY 1501 FVTPGERSGVFDSSVLCCEYDAGCAWYELTFAETSVRILRAYLNTPGIPVQODHLEFVES 1560
Db 1501 FVTPGERSGVFDSSVLCCEYDAGCAWYELTFAETSVRILRAYLNTPGIPVQODHLEFVES 1560
QY 1561 VFTGLTHIDAIFLSQTKQAGDNFVYLVAIQATVQCARAAPPSPSWDMWKCLIRLKTPLHG 1620
Db 1561 VFTGLTHIDAIFLSQTKQAGDNFVYLVAIQATVQCARAAPPSPSWDMWKCLIRLKTPLHG 1620
QY 1621 PTPLLYRLGAVQNEVILCHPTIKYIMACMSADLEVVVTSVWLVGGVLAALAAAYCLTITGSV 1680
Db 1621 PTPLLYRLGAVQNEVILCHPTIKYIMACMSADLEVVVTSVWLVGGVLAALAAAYCLTITGSV 1680
QY 1681 VIVGRIILSGRPANVPDREVLYQBFDEMBECASQIPYTEQGMOLABOPKOKALGILQAT 1740
Db 1681 VIVGRIILSGRPANVPDREVLYQBFDEMBECASQIPYTEQGMOLABOPKOKALGILQAT 1740
QY 1741 KOEAAAPVVBESKWRLETFWAKEMNFIISGIOYLGLSTLPGNPAIASLMAFTASITSP 1800
Db 1741 KOEAAAPVVBESKWRLETFWAKEMNFIISGIOYLGLSTLPGNPAIASLMAFTASITSP 1800
QY 1801 LTTQNTLLFNILGCVAAQLAPPSSAAGVAGIAGAAVSGISGLGKVLVDILAGYGAGVA 1860
Db 1801 LTTQNTLLFNILGCVAAQLAPPSSAAGVAGIAGAAVSGISGLGKVLVDILAGYGAGVA 1860
QY 1861 GALVAFKVMSEVESTEDLVNLLPAILSPGALVGVVCAALIRRHVGPGEAVOMMELI 1920
Db 1861 GALVAFKVMSEVESTEDLVNLLPAILSPGALVGVVCAALIRRHVGPGEAVOMMELI 1920
QY 1921 AFASRGHNVSPTHYVPESDAAARVTQILSSLTITQLLKRHLQWINEDECSTPCSCSWLRDV 1980
Db 1921 AFASRGHNVSPTHYVPESDAAARVTQILSSLTITQLLKRHLQWINEDECSTPCSCSWLRDV 1980
QY 1981 WDMTCTVLTDFTKVLQSKLLPRLPGVFLSCORQYKGVWRGDMQMTCPGAGIAGHVK 2040
Db 1981 WDMTCTVLTDFTKVLQSKLLPRLPGVFLSCORQYKGVWRGDMQMTCPGAGIAGHVK 2040
QY 2041 NGSMRIIVGPRTCSTWHTGTPINAYTTGPTCPSPAPNSRALMRVAEEVYVTRVGDHF 2100
Db 2041 NGSMRIIVGPRTCSTWHTGTPINAYTTGPTCPSPAPNSRALMRVAEEVYVTRVGDHF 2100
QY 2101 YVTCMTTNDNVKCPQVPAPEFFTEVDGVRHLRYAPACKPLLEDVTPQVGLNQVLVSQ 2160
Db 2101 YVTCMTTNDNVKCPQVPAPEFFTEVDGVRHLRYAPACKPLLEDVTPQVGLNQVLVSQ 2160
QY 2161 PCEPEPDVTLTSLMTPDPSHITAEAKRLARQSPPSLASSASQLSAPSLKATCTTHHD 2220
Db 2161 PCEPEPDVTLTSLMTPDPSHITAEAKRLARQSPPSLASSASQLSAPSLKATCTTHHD 2220
QY 2221 SPDADLLEANLLWPCQMGMITRVESENKVVILDSPEPLHABGDEREISVAAIRLRSKR 2280
Db 2221 SPDADLLEANLLWPCQMGMITRVESENKVVILDSPEPLHABGDEREISVAAIRLRSKR 2280
QY 2281 PPSALPTIWARPDYNNPPLLESWKOPDYVPPVPHGCPPTPKAPPIPPPRKRTVVLTESV 2340
Db 2281 PPSALPTIWARPDYNNPPLLESWKOPDYVPPVPHGCPPTPKAPPIPPPRKRTVVLTESV 2340
QY 2341 SSALAEATKTFGSSGSSAVDSGTATALPLADDDGDKGSDVESYSMPLEGEPPGDD 2400
Db 2341 SSALAEATKTFGSSGSSAVDSGTATALPLADDDGDKGSDVESYSMPLEGEPPGDD 2400
QY 2401 SDGWSVTSSEASEDVVCCSMSTWTCALITPCAABESKLPINPLNSLRHHNVYATT 2460
Db 2401 SDGWSVTSSEASEDVVCCSMSTWTCALITPCAABESKLPINPLNSLRHHNVYATT 2460
QY 2461 SRASLRQKVTDFRDLQVLDHVDVLEKMKAKASTVKAKLLSIEEACKLTTPHSAZSKF 2520

Db 2461 SRASLRQKVTDFRLQVLDHRYDVLKEMKAKASTVKAKLLSVEEACKLTPPHSARSKF 2520
 Qy 2521 GYGAQVNLSSRAVNHRSVWDLLETPIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
 Db 2521 GYGAQVNLSSRAVNHRSVWDLLETPIDTTIMAKSEVFCVQPEKGRKPARLIV 2580
 Qy 2581 FPDGLGVRCERKALYDVVSTLPOAVMGSSYGFQSPKQRFVFLVNTWKSKECPMGFSYDT 2640
 Db 2581 FPDGLGVRCERKALYDVVSTLPOAVMGSSYGFQSPKQRFVFLVNTWKSKECPMGFSYDT 2640
 Qy 2641 RCFDSTVIESIRVEESIQCCDLAPEARQAIRSLTERLYTGGPLTNSKGQNGYRCRA 2700
 Db 2641 RCFDSTVIESIRVEESIQCCDLAPEARQAIRSLTERLYTGGPLTNSKGQNGYRCRA 2700
 Qy 2701 SGVLTTSCTGNTLCYVKAKAARAKLODCTMLVNGDDLWVICESAGTOEDAAALRAFTE 2760
 Db 2701 SGVLTTSCTGNTLCYVKAKAARAKLODCTMLVNGDDLWVICESAGTOEDAAALRAFTE 2760
 Qy 2761 AMTRYSAAPGPPPEYDLELITSCSSNVSAHDASGKRVVYLTTRDPTTPLARAWEITAR 2820
 Db 2761 AMTRYSAAPGPPPEYDLELITSCSSNVSAHDASGKRVVYLTTRDPTTPLARAWEITAR 2820
 Qy 2821 HTPINSWLGNIIMVAPTLMWAMWILMTHEFSILLAQEQLKALDCQIYGACYSIEPLDLPQ 2880
 Db 2821 HTPINSWLGNIIMVAPTLMWAMWILMTHEFSILLAQEQLKALDCQIYGACYSIEPLDLPQ 2880
 Qy 2881 IIERLHGLSAFTLSYSGEINRVASCLRLKLGVPPLRTWRHARSVRAKLSQCGRAATC 2940
 Db 2881 IIERLHGLSAFTLSYSGEINRVASCLRLKLGVPPLRTWRHARSVRAKLSQCGRAATC 2940
 Qy 2941 GRYLFNNAVRKILKLTTPAASQDLSCWFFVAGYSGGDIYHSLGRAPRPWPLCLLLLSV 3000
 Db 2941 GRYLFNNAVRKILKLTTPAASQDLSCWFFVAGYSGGDIYHSLGRAPRPWPLCLLLLSV 3000
 Qy 3001 GVGIIYLLPNR 3010
 Db 3001 GVGIIYLLPNR 3010

RESULT 10
 ID ABG32461 standard; protein; 3010 AA.
 XX AC ABG32461;
 XX DT 15-NOV-2002 (first entry)
 XX DE Hepatitis C virus Con 1 isolate polyprotein mutant #10.
 XX KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; muten;
 XX KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
 XX KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
 XX OS Hepatitis C virus.
 XX OS Synthetic.
 XX FH Location/Qualifiers
 FT MISC-difference 2204
 FT /note= "Wild-type Ser substituted by Arg"
 XX WO2000259321-A2.
 XX 01-AUG-2002.
 XX 16-JAN-2002; 2002WO-EP000526.
 XX 23-JAN-2001; 2001US-0263479P.
 XX (RICE-) IST RICECHE BIOL MOLECOLARE ANGELETTI.
 XX De Francesco R, Migliaccio G, Paonessa G;
 XX WPI; 2002-599793/64.

XX New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
 PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
 PT ribosome entry site (IRES) region, useful in studying HCV replication and
 PT expression.
 XX Claim 1; Page: 69pp; English.
 XX The invention relates to nucleic acid molecules comprising altered HCV
 CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
 CC internal ribosome entry site (IRES) region coding for one or more NS3,
 CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
 CC are detailed in the specification. Also included are (1) an expression
 CC vector comprising a nucleotide sequence coding for the altered nucleic
 CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
 CC recombinant cell human hepatoma cell comprising the altered nucleic acids
 CC ; (3) a recombinant cell produced by introducing into a human hepatoma
 CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
 CC replicon enhanced cell or which containing a functional HCV replicon; (5)
 CC an HCV replicon enhanced cells made in the method; and (6) measuring the
 CC ability of a compound to affect HCV activity. The HCV replicons and HCV
 CC replicon enhanced cells are useful in studying HCV replication and
 CC expression, and HCV and host cell interactions, producing HCV RNA and
 CC proteins, and providing a system for measuring the ability of a compound
 CC to modulate one or more HCV activities e.g. to discover drugs which may
 CC treat HCV mediated diseases such as liver failure, cirrhosis and
 CC hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
 CC polyprotein (comprising the Core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A
 CC and NS5B proteins), NS5A mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the HCV sequence appearing as ABG32451 and the information in claim
 CC 1
 XX Sequence 3010 AA;
 SQ Query Match 95.7%; Score 15318; DB 5; Length 3010;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 2851; Conservative 76; Mismatches 83; Indels 0; Gaps 0;
 Qy 1 MSTNPKPQKTKENTNRPODVKPRGGQIVGGVYLLPRGPELGVRAATKASERQPRG 60
 Db 1 MSTNPKPQKTKENTNRPODVKPRGGQIVGGVYLLPRGPELGVRAATKASERQPRG 60
 Qy 61 RRQPIKARPEGRANWQPCYPWPLXGNEGLWAGWLLSPRGSRPSWGTPDPRRRNLG 120
 Db 61 RRQPIKARPEGRANWQPCYPWPLXGNEGLWAGWLLSPRGSRPSWGTPDPRRRNLG 120
 Qy 121 KVDTTTCGPADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
 Db 121 KVDTTTCGPADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
 Qy 181 LLSCLTIPASAYEVRNVSGIYHYVTNDCNSSIIVEAADVIMHTPGCVPCVOECNSRRCW 240
 Db 181 LLSCLTIPASAYEVRNVSGIYHYVTNDCNSSIIVEAADVIMHTPGCVPCVOECNSRRCW 240
 Qy 241 ALPTTLAARNASVPTTIRRHVDLLVGTAAFCSAMVYVGLDLOGSIFLVSLFTSPRHER 300
 Db 241 ALPTTLAARNASVPTTIRRHVDLLVGTAAFCSAMVYVGLDLOGSIFLVSLFTSPRHER 300
 Qy 301 VQDCNCSIYPCHVSGHRMAMDMNMNSPTALVVSQILLRIPQAVVDMVAGAHGVLAGLA 360
 Db 301 VQDCNCSIYPCHVSGHRMAMDMNMNSPTALVVSQILLRIPQAVVDMVAGAHGVLAGLA 360
 Qy 361 YYSWVGWNAKVLIVALLFAGVDGETHTTGRVAGHTTSGFTSLFSSGASOKIQLVNTNGSW 420
 Db 361 YYSWVGWNAKVLIVALLFAGVDGETHTTGRVAGHTTSGFTSLFSSGASOKIQLVNTNGSW 420
 Qy 421 HINRTALNCNDSLQGTGFFAALFYAHKFNSSGCGPERMASCRPIDWFAQGWGPTITYPNSS 480
 Db 421 HINRTALNCNDSLQGTGFFAALFYAHKFNSSGCGPERMASCRPIDWFAQGWGPTITYPNSS 480
 Qy 481 DORPYCNHAYAPRCGVVPASQVCGPVYCTPTSPVVTGTTDRSGVPTYSWGENETDNLN 540

481 DORPYCWHYAPRPGIVPAQVCGPVYCTFSPVVGTTDRFVPTYSWGENETDVLIN 540
491 NTRPPQGNWPGCTWMSSTGFTKCGPPNCGVGNRTLICPTDCFRKHPAEATYKCGSG 600
501 NTRPPQGNWPGCTWMSSTGFTKCGPPNCGVGNRTLICPTDCFRKHPAEATYKCGSG 600
601 PWLTBRCLVDYPRVLRHYPCTNFSTFKVBYMYGVGVEHRLNAACNTRGRRCNLEDRDRS 660
601 PWLTBRCLVDYPRVLRHYPCTNFSTFKVBYMYGVGVEHRLNAACNTRGRRCNLEDRDRS 660
661 ELSPLLLSTTEWOLPCAFPTLLPALSTGLHLHONTVDVQVLYXGVSFAFSAIKWEYIL 720
661 ELSPLLLSTTEWOLPCAFPTLLPALSTGLHLHONTVDVQVLYXGVSFAFSAIKWEYIL 720
721 LFLLLADARVCACLWMLLIIAQAEALENLVNLNAASVAGAHGILSLFVFFCAAWYIKG 780
721 LFLLLADARVCACLWMLLIIAQAEALENLVNLNAASVAGAHGILSLFVFFCAAWYIKG 780
781 RLAPGAAAYFYGVWPLLLIALLPRAAYALDREMAASCGGAVLVGLVFLTSLPYKVLFT 840
781 RLVPFGAAAYLYGVWPLLLIALLPRAAYALDREMAASCGGAVLVGLVFLTSLPYKVLFT 840
841 RLITWMLQYFTRAEAHQVWVPLNVRGGRDAIILLTCAVHPBLIFDITKLLAILGPLM 900
841 RLITWMLQYFTRAEAHQVWVPLNVRGGRDAIILLTCAVHPBLIFDITKLLAILGPLM 900
901 VLQAGITRVYFVRAAGLIRACMLVRKAVAGHTVQVFWKLGALGTGVYVNHLPJRDWA 960
901 VLQAGITRVYFVRAAGLIRACMLVRKAVAGHTVQVFWKLGALGTGVYVNHLPJRDWA 960
961 HAQLRLDVAVAPVVFVSAMETKVITWGADTAACGDIIILGLPVSARGKEIFLGPADSLRG 1020
961 HAQLRLDVAVAPVVFVSAMETKVITWGADTAACGDIIILGLPVSARGKEIFLGPADSLRG 1020
1021 QWRLLIAPITAYSQOQTRGVLCIIITSLTGRDNQVGEVQVWSTAQSFALATCINGCVMT 1080
1021 QWRLLIAPITAYSQOQTRGVLCIIITSLTGRDNQVGEVQVWSTAQSFALATCINGCVMT 1080
1081 VYHGAGSKTLAPKGPITOMYTNVDLDLVGWAQPPGARNTPCSSDLYLVTRHADYI 1140
1081 VYHGAGSKTLAPKGPITOMYTNVDLDLVGWAQPPGARNTPCSSDLYLVTRHADYI 1140
1141 PYRRGDSRGLLSRPPVSYLYGSSGGPLLCPSGHVGVFRAVCTRGYAKAVDTPVPS 1200
1141 PYRRGDSRGLLSRPPVSYLYGSSGGPLLCPSGHVGVFRAVCTRGYAKAVDTPVPS 1200
1201 METTMRSPVFTDNSTPPAVPQTFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAA 1260
1201 METTMRSPVFTDNSTPPAVPQTFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAA 1260
1261 TLGFGAYMSKAGHDNPNTGTGRTTITGSIITYSYGKFLADGGCGGGAYDIIICDECHS 1320
1261 TLGFGAYMSKAGHDNPNTGTGRTTITGSIITYSYGKFLADGGCGGGAYDIIICDECHS 1320
1321 TDSTTILGTVLDQAEATAGARLVLATATPGSVTVPHNTPEETGLSNNGEIPYKAI 1380
1321 TDSTTILGTVLDQAEATAGARLVLATATPGSVTVPHNTPEETGLSNNGEIPYKAI 1380
1381 PIRAIKGRHLIFCHSKKCDLAKLTGLGLNAVAYYRGDVSVIPPIGDVWVATDAL 1440
1381 PIRAIKGRHLIFCHSKKCDLAKLTGLGLNAVAYYRGDVSVIPPIGDVWVATDAL 1440
1441 MTGFTGDSVTDKNTCVTQVDFSLDPTFTTETTPQDAVRSQRGRGTGRSGIYR 1500
1441 MTGFTGDSVTDKNTCVTQVDFSLDPTFTTETTPQDAVRSQRGRGTGRSGIYR 1500
1501 FVTPGERPSGMPDSSVLCEDYDAGCANVELTIPAEITSVRLRAYLNTPGVPVODHLRFWES 1560
1501 FVTPGERPSGMPDSSVLCEDYDAGCANVELTIPAEITSVRLRAYLNTPGVPVODHLRFWES 1560
1561 VFTGLTHIDAHPLSOTKQAGDNFPYLVAQVCAQAQPPPSWDMKCLIRLXPTLHG 1620
1561 VFTGLTHIDAHPLSOTKQAGDNFPYLVAQVCAQAQPPPSWDMKCLIRLXPTLHG 1620

1621 PTELLYRLGAVQNEVILTHPIKTYIMACWSADLEVVVTSTWVLGGVLAALAAVCLTTGVS 1680
1621 PTELLYRLGAVQNEVILTHPIKTYIMACWSADLEVVVTSTWVLGGVLAALAAVCLTTGVS 1680
1681 VIVGRIILSGKPAVVDREVLYQOEDEMEECASQLPYIBQGMQLAPQFKQKALGLLQAT 1740
1681 VIVGRIILSGKPAVVDREVLYQOEDEMEECASQLPYIBQGMQLAPQFKQKALGLLQAT 1740
1741 KQAEAAAPVVEKSWALETFWAKMKNFISGTOYLAGISTLPCNPAIASLMAFTASITSP 1800
1741 KQAEAAAPVVEKSWALETFWAKMKNFISGTOYLAGISTLPCNPAIASLMAFTASITSP 1800
1801 LITQNTLLFNILGGWVAQAALAPPSAASAPVAGIAGAAVSGTGLGKVLVDIILAGTGAGVA 1860
1801 LITQNTLLFNILGGWVAQAALAPPSAASAPVAGIAGAAVSGTGLGKVLVDIILAGTGAGVA 1860
1861 GALVAFKVMGVBVSTEDLVNLLPALISPGALVVGWVCAAILRRHVGPGEAGVQWNRLLI 1920
1861 GALVAFKVMGVBVSTEDLVNLLPALISPGALVVGWVCAAILRRHVGPGEAGVQWNRLLI 1920
1921 APASRGNHVSPTHYVPESDAAARVTQIILSSLTITOLIKELHOWINEDCSTPCSGSWLRDV 1980
1921 APASRGNHVSPTHYVPESDAAARVTQIILSSLTITOLIKELHOWINEDCSTPCSGSWLRDV 1980
1981 WDWICTVLTDFTXTLQSKLLPRLPGVPFLSCQRYGVWVRGDIQMTTCCPQAGIAGHVK 2040
1981 WDWICTVLTDFTXTLQSKLLPRLPGVPFLSCQRYGVWVRGDIQMTTCCPQAGIAGHVK 2040
2041 NGSMBIVGPRICNTWHTGTFPINAYTTGCTPSPAPNTSRALNRVAASEYVEVTRVGFPH 2100
2041 NGSMBIVGPRICNTWHTGTFPINAYTTGCTPSPAPNTSRALNRVAASEYVEVTRVGFPH 2100
2101 VYTGMTTNVKKPCQVPAPEPTEVDGVLHRYAPACKPLLRREDVTFOVGLNQYLVGSOL 2160
2101 VYTGMTTNVKKPCQVPAPEPTEVDGVLHRYAPACKPLLRREDVTFOVGLNQYLVGSOL 2160
2161 PCEPDPVTLTSMLTDPSHITAEAKRLARGSPPSLASSASOLSAAPSILKATCTTHD 2220
2161 PCEPDPVTLTSMLTDPSHITAEAKRLARGSPPSLASSASOLSAAPSILKATCTTHD 2220
2221 SPDADLIENLNRQEMGNITRVSEKNVILDSFEPHAEGEREISVAABILRKSRK 2280
2221 SPDADLIENLNRQEMGNITRVSEKNVILDSFEPHAEGEREISVAABILRKSRK 2280
2281 PPSALPIWARPDYNPILLESKDPDYPVPHGCPPLTPKAPPPIPPPRKRTVVLTSNV 2340
2281 PPSALPIWARPDYNPILLESKDPDYPVPHGCPPLTPKAPPPIPPPRKRTVVLTSNV 2340
2341 SSALAEALATKTFGSGSSAVDSGTATAPDQPSDDGAGSDVESYSSMPPLEGPDPDL 2400
2341 SSALAEALATKTFGSGSSAVDSGTATAPDQPSDDGAGSDVESYSSMPPLEGPDPDL 2400
2401 SDGSKSTVSEASRDNVCCSMSTWTGALITPCAAEESKLPIINPLSNLLRHHNVAT 2460
2401 SDGSKSTVSEASRDNVCCSMSTWTGALITPCAAEESKLPIINPLSNLLRHHNVAT 2460
2461 SRSASLRQKVTFFDLQVLDHRYDVLKEMKAKASTVAKILSTEEACKLTPPSAKSKF 2520
2461 SRSASLRQKVTFFDLQVLDHRYDVLKEMKAKASTVAKILSTEEACKLTPPSAKSKF 2520
2521 CYGAKDVNLSRANVHRSVWEDLEDETETFTITIMAKSEVFCVQPEKGRKPARLIV 2580
2521 CYGAKDVNLSRANVHRSVWEDLEDETETFTITIMAKSEVFCVQPEKGRKPARLIV 2580
2581 FPDGLVAVCEKMALDYVSTILPQAVNGSSYGFQYSPQRVFLVNTWKSXKCPMGFSYDT 2640
2581 FPDGLVAVCEKMALDYVSTILPQAVNGSSYGFQYSPQRVFLVNTWKSXKCPMGFSYDT 2640
2641 RCFDSTVDESIRVEESYQCCDLAPEARQAIRSLTERLYIGGELTNSKQNCYRRCRA 2700
2641 RCFDSTVDESIRVEESYQCCDLAPEARQAIRSLTERLYIGGELTNSKQNCYRRCRA 2700

QY 2701 SGVLTSCGNTLCYKATACAAKLOCTMIVNGEDLVVICESAGTQEDAAALRAFTE 2760
Db SGVLTSCGNTLCYKATACAAKLOCTMIVNGEDLVVICESAGTQEDAAALRAFTE 2760
QY 2761 AMTRYSAPGDPDPQPYDLELITSCSNYSVAHDASGKVVYLTRDPTTFLARAWEATAR 2820
Db AMTRYSAPGDPDPQPYDLELITSCSNYSVAHDASGKVVYLTRDPTTFLARAWEATAR 2820
QY 2821 HTPINSWLGNIIMYAFILWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPDLPO 2880
Db HTPINSWLGNIIMYAFILWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPDLPO 2880
QY 2881 ITERLHGLSFTHSYSPGEINRVASCLXIGVPPPLRWHRARSVRKLLSQGGRATC 2940
Db ITERLHGLSFTHSYSPGEINRVASCLXIGVPPPLRWHRARSVRKLLSQGGRATC 2940
QY 2941 GRYLENWAVETKLTPIPAASQDLSSWVAGYSGDDIYHSISRARPEWFLCILLISV 3000
Db GRYLENWAVETKLTPIPAASQDLSSWVAGYSGDDIYHSISRARPEWFLCILLISV 3000
QY 3001 GVGIVILLPNR 3010
Db GVGIVILLPNR 3010

RESULT 11

ID ABG32458 standard; protein; 3010 AA.
AC ABG32458;
D7 15-NOV-2002 (first entry)

DE Hepatitis C virus Con 1 isolate polyprotein mutant #7.

KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; muten;
KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
XX Hepatitis C virus.
OS Synthetic.

Key Location/Qualifiers

FT Misc-difference 2197 /note= "Wild-type Phe substituted by Ser"

XX WO200259321-A2.

PD 01-AUG-2002.

XX 16-JAN-2002; 2002WO-EP000526.

XX 23-JAN-2001; 2001US-0263479P.

PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX De Francesco R, Migliaccio G, Paonessa G;

XX WPI, 2002-599793/64.

PT New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT ribosome entry site (IRES) region, useful in studying HCV replication and
PT expression.
XX Claim 1; Page; 69pp; English.

XX The invention relates to nucleic acid molecules comprising altered HCV
CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC internal ribosome entry site (IRES) region coding for one or more NS3,
CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC are detailed in the specification. Also included are (1) an expression
CC vector comprising a nucleotide sequence coding for the altered nucleic

acids, which is transcriptionally coupled to an exogenous promoter; (2) a
recombinant cell human hepatoma cell comprising the altered nucleic acids
; (3) a recombinant cell produced by introducing into a human hepatoma
cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
replicon enhanced cell or which containing a functional HCV replicon; (5)
an HCV replicon enhanced cells made in the method; and (6) measuring the
ability of a compound to affect HCV activity. The HCV replicons and HCV
replicon enhanced cells are useful in studying HCV replication and
expression, and HCV and host cell interactions, producing HCV RNA and
proteins, and providing a system for measuring the ability of a compound
to modulate one or more HCV activities e.g. to discover drugs which may
treat HCV mediated diseases such as liver failure, cirrhosis and
hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
polyprotein (comprising the Core, E1, E2, E7, NS2, NS3, NS4A, NS4B, NS5A
and NS5B proteins), NS5A mutant of the invention. Note: The present
sequence is not shown in the specification but was created by the indexer
using the HCV sequence appearing as ABG32451 and the information in claim
1

XX Sequence 3010 AA;

Query Match 95.7%; Score 15317; DB 5; Length 3010;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 2851; Conservative 76; Mismatches 83; Indels 0; Gaps 0;

QY 1 MSTNPKPQRTKRNTRRNPQDVXZPPGGQIVGGVYLLPRGPRGLGVRATKASRSOPRG 60
Db 1 MSTNPKPQRTKRNTRRNPQDVXZPPGGQIVGGVYLLPRGPRGLGVRATKASRSOPRG 60

QY 61 RQPIPKARQEGRAWAOPGYEWPPLYGNEGILGAGWLLSPRSGRSPSGPTDPRRSRNLG 120
Db 61 RQPIPKARQEGRAWAOPGYEWPPLYGNEGILGAGWLLSPRSGRSPSGPTDPRRSRNLG 120

QY 121 KVIDTLACGADLMGYIPLVCAPLGGAARALAGVRVLEEDGVNATGNLPGCSFSIFLLA 180
Db 121 KVIDTLACGADLMGYIPLVCAPLGGAARALAGVRVLEEDGVNATGNLPGCSFSIFLLA 180

QY 181 LLSCLTIPASAYEVNRVNSGIYVHTNDCSNSIIVYEAADVIHMTDPCVPCVQEGNSSRCW 240
Db 181 LLSCLTIPASAYEVNRVNSGIYVHTNDCSNSIIVYEAADVIHMTDPCVPCVQEGNSSRCW 240

QY 241 ALTTLAARNASVPTTIRRHVDLLVGTAAFCASAYVGLCGSIFLVSQLFTSPRHHET 300
Db 241 ALTTLAARNASVPTTIRRHVDLLVGTAAFCASAYVGLCGSIFLVSQLFTSPRHHET 300

QY 301 VQDCNCSIYPGHVSGHEWANDMMNWSPTTALYVSQLRIPQAVVDMVAGAHGVLGAGLA 360
Db 301 VQDCNCSIYPGHVSGHEWANDMMNWSPTTALYVSQLRIPQAVVDMVAGAHGVLGAGLA 360

QY 361 YISMVGNWAKVLIIVALLFAGVDGRTHTGRVAGHTTSGFTSLFSSGASOKIQLVNTNGSW 420
Db 361 YISMVGNWAKVLIIVALLFAGVDGRTHTGRVAGHTTSGFTSLFSSGASOKIQLVNTNGSW 420

QY 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCCPERMASCPIDNFAQGWGPIYTKPNSS 480
Db 421 HINRTALNCNDSLOTGFFAALFYAHKFNSSGCCPERMASCPIDNFAQGWGPIYTKPNSS 480

QY 481 DQRPYCWHEYAPRCGVVPASQVGVVCTPSPVAVGTGTDSDRGVPTYSWGENETDVLIN 540
Db 481 DQRPYCWHEYAPRCGVVPASQVGVVCTPSPVAVGTGTDSDRGVPTYSWGENETDVLIN 540

QY 541 NTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGNRTLICPTDCFRKHPAATYTKCGSG 600
Db 541 NTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGNRTLICPTDCFRKHPAATYTKCGSG 600

QY 601 PWLTPLCLVDYPRALWHYPCTLNPSIFKVMYVGVGVEHRLNAAKNTRGERCNLEDRDS 660
Db 601 PWLTPLCLVDYPRALWHYPCTLNPSIFKVMYVGVGVEHRLNAAKNTRGERCNLEDRDS 660

QY 661 ELSPLILSTTEWOLPCAFITLTPALSTGLHFLHONIVDQYLYGVGSFAVFSFAIKWEYVL 720
Db 661 ELSPLILSTTEWOLPCAFITLTPALSTGLHFLHONIVDQYLYGVGSFAVFSFAIKWEYVL 720

QY 721 LFLALLADARVACLAWMMLIAQAAEALENLVVINAASVAGARGILSFLVFFCAAWYIKG 780
Db 721 LFLALLADARVACLAWMMLIAQAAEALENLVVINAASVAGARGILSFLVFFCAAWYIKG 780
QY 781 RLAPGAAYAVGVWPULLLLALPPRAYALDREMAASCGGAVLVGLVFTLSPPYKVFLLT 840
Db 781 RLVPGAAYAVGVWPULLLLALPPRAYALDREMAASCGGAVLVGLVFTLSPPYKVFLLT 840
QY 841 RLILWLQYFTTRAHAHQWVWPPLNVRGGRDALILITCAVHPELIPDITKLLAILGLPM 900
Db 841 RLILWLQYFTTRAHAHQWVWPPLNVRGGRDALILITCAVHPELIPDITKLLAILGLPM 900
QY 901 VLQAGITRPYFVRAOGLIRACMLVRKAGGHVQVMVFKLGAITGTYVYNHLTPLRDWA 960
Db 901 VLQAGITKVPYFVRAHGLIRACMLVRKAGGHVQVMALMKLAALTGTYYVYDHLTPLRDWA 960
QY 961 HAGLRDLAVAVEPVFSAMETKITWGAADTAACGDIIILGLPVSAARGKEIFLGPADSLEG 1020
Db 961 HAGLRDLAVAVEPVFSAMETKITWGAADTAACGDIIILGLPVSAARGKEIFLGPADSLEG 1020
QY 1021 QGWRLLAPITAYSQOTRGVLGCIITSIGTRDKQVGEVQVVSSTATQSFATCINGVCWT 1080
Db 1021 QGWRLLAPITAYSQOTRGVLGCIITSIGTRDRNQVEGEVQVVSSTATQSFATCINGVCWT 1080
QY 1081 VYHGAGSKTLAGPKGPIITQMYTNVDLDLVGWQAPPGARSWTPCSCSSDLYIVTRHADYV 1140
Db 1081 VYHGAGSKTLAGPKGPIITQMYTNVDLDLVGWQAPPGARSLTPCTCGSSDLYIVTRHADYV 1140
QY 1141 PYERRGDSRGSLLSPRPVSYLKSSGGPILCPSGHVGVFRAAVCTRGYAKAVDFIPVES 1200
Db 1141 PYERRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGYAKAVDFIPVES 1200
QY 1201 METTMRSPVFTDNSTPPAVPOTQVAHLHAPTGSGLTAKVPAAYAAQGYKVLVLPNSVAA 1260
Db 1201 METTMRSPVFTDNSTPPAVPOTQVAHLHAPTGSGLTAKVPAAYAAQGYKVLVLPNSVAA 1260
QY 1261 TLGFGHVMKARGIDNIBTGRTITGGSITYSTYKFIAGCGCGGAYDIIICDECHS 1320
Db 1261 TLGFGHVMKARGIDNIBTGRTITGAPITVSTYKFIAGCGCGGAYDIIICDECHS 1320
QY 1321 TDSTTILGTGLVLDQETAGALVLVLTATPPGSSVTVPHENITEETGLSNNGIPFGKAI 1380
Db 1321 TDSTTILGTGLVLDQETAGALVLVLTATPPGSSVTVPHENITEETGLSNNGIPFGKAI 1380
QY 1381 PIEALTKGHHILFCHSKKXKDBELAKLGLGLNAVAYRGLDVSVIPPIGDVVVATDAL 1440
Db 1381 PIETTKGRRHLIFCHSKKXKDBELAKLGLGLNAVAYRGLDVSVIPPIGDVVVATDAL 1440
QY 1441 MTGFTGDFSDVDCNCTVQTQVDFSLDPTFTIETTVPODAVSRQRRGTGRGSIYR 1500
Db 1441 MTGFTGDFSDVDCNCTVQTQVDFSLDPTFTIETTVPODAVSRQRRGTGRGSIYR 1500
QY 1501 FVTPGERSGMFDSSVLCYDAGCAWYBELTFAETSVRILRAYLNTPLGLPVCODHLEFWS 1560
Db 1501 FVTPGERSGMFDSSVLCYDAGCAWYBELTFAETSVRILRAYLNTPLGLPVCODHLEFWS 1560
QY 1561 VFTGLTHIDAHFLSOTKQAGDNFPYLVAQATVCARAQAPPSWDQMKCLRLKPTLHG 1620
Db 1561 VFTGLTHIDAHFLSOTKQAGDNFPYLVAQATVCARAQAPPSWDQMKCLRLKPTLHG 1620
QY 1621 PTELLYRLGAVONEVLLTHPIITKYIMACNSADLEVVTSTWLVGGVLAALAYCLITGSV 1680
Db 1621 PTELLYRLGAVONEVTTTHPIITKYIMACNSADLEVVTSTWLVGGVLAALAYCJTGSV 1680
QY 1681 VIVGRILSGKPAVDPDREVLVYOEFDMEECASQLPYIBQGMQLAQFQKQALGLLQAT 1740
Db 1681 VIVGRILSGKPAIPDREVLVYOEFDMEECASHLYPIBQGMQLAQFQKQALGLLQAT 1740
QY 1741 KOAEAAAAPVVEKWALETFWAKHWNFTSGIOYLAGLSTLPGNPAIASLMAFTASITSP 1800
Db 1741 KOAEAAAAPVVEKWRLEAFWAKHWNFTSGIOYLAGLSTLPGNPAIASLMAFTASITSP 1800
QY 1801 LTTQNTLLFNILGGWVAQAALAPPSAASAFVAGIAGAAVGSIGLKVLDIILAGVAGVA 1860

Db 1801 LTTQNTLLFNILGGWVAQAALAPPSAASAFVAGIAGAAVGSIGLKVLDIILAGVAGVA 1860
QY 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVWCIAILRRHVGPGEVAVQWNRLLI 1920
Db 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVWCIAILRRHVGPGEVAVQWNRLLI 1920
QY 1921 AFASRGNHVSPTHYPESDAAARVTQILSSSIFITOLLKXHOWINEDSTPCSGSWLRDV 1980
Db 1921 AFASRGNHVSPTHYPESDAAARVTQILSSSIFITOLLKXHOWINEDSTPCSGSWLRDV 1980
QY 1981 WDWICTVLTDFTXMTLQSKLLPRLPGVPFLSCQRGKGYKWRGDIQCTTCCPQAGIAGHVK 2040
Db 1981 WDWICTVLTDFTXMTLQSKLLPRLPGVPFLSCQRGKGYKWRGDIQCTTCCPQAGIAGHVK 2040
QY 2041 NGSMLIVGPRTCNSNTHGTFPFINAYTTGCTPSPAPNYSRALWRVAABEYVEVTRVGFH 2100
Db 2041 NGSMLIVGPRTCNSNTHGTFPFINAYTTGCTPSPAPNYSRALWRVAABEYVEVTRVGFH 2100
QY 2101 VYTGMTNDNVKPCQVPAPEFTEVDGVLHRYAPACKPLLEDVTFVGLNQYLVSQOL 2160
Db 2101 VYTGMTNDNVKPCQVPAPEFTEVDGVLHRYAPACKPLLEDVTFVGLNQYLVSQOL 2160
QY 2161 PCEPEPDVTVLTSMLTDPSTHITAEAKRLARGSPSPASSASQLSAPSLKATCTTHD 2220
Db 2161 PCEPEPDVAVLTSMLTDPSTHITAEAKRLARGSPSPASSASQLSAPSLKATCTTHD 2220
QY 2221 SPDAOLIIENLLWROEMGNITRVESENKVILDSFEPLHAEGBDEETISVAABILRRSK 2280
Db 2221 SPDAOLIIENLLWROEMGNITRVESENKVILDSFEPLHAEGBDEETISVAABILRRSK 2280
QY 2281 FPSALPIWARPDYNPPLLESWKDPDYVPVVGCPPLPTKAPPIPPRRKRTVWLTESNV 2340
Db 2281 FPRAMPINARPDYNPPLLESWKDPDYVPVVGCPPLPAKAPPIDPPRRKRTVWLTESNV 2340
QY 2341 SSALAEALATKTFGSGSSAVDSGTATAPDLSADDDGKGSVDVESYSSMPPLEGEPDPL 2400
Db 2341 SSALAEALATKTFGSGSSAVDSGTATAPDLSADDDGKGSVDVESYSSMPPLEGEPDPL 2400
QY 2401 SDGWSVTSEASESDVVCSSSYTWTGALITPCAABESKLINLNSLLRHNXYATT 2460
Db 2401 SDGWSVTSEASESDVVCSSSYTWTGALITPCAABETKLINLNSLLRHNXYATT 2460
QY 2461 SRSASLRQKVTDFDLQVLDHRYDVLKEMKAKASTVAKALLSTBEACKLPPHISAKKF 2520
Db 2461 SRSASLRQKVTDFDLQVLDHRYDVLKEMKAKASTVAKALLSTBEACKLPPHISAKKF 2520
QY 2521 GYGAKDVNLSSRAVNHIRSVWDLLEDTEPITDITIMAKSEVFCVQPEKGRKPARLIV 2580
Db 2521 GYGAKDVNLSSRAVNHIRSVWDLLEDTEPITDITIMAKNEVFCVQPEKGRKPARLIV 2580
QY 2581 FPD-LGVRVCEKMAVYDVVSTLPOAVMGSSYGFQYSPKORVZFLNVTWKSCKPCMFSDY 2640
Db 2581 FPD-LGVRVCEKMAVYDVVSTLPOAVMGSSYGFQYSPKORVZFLNVTWKSCKPCMFSDY 2640
QY 2641 RCFDSTVTESDIRVEESIYQCCDLAPARQAIRSILTERLYTGGPLTNSKQNGVRRORA 2700
Db 2641 RCFDSTVTESDIRVEESIYQCCDLAPARQAIRSILTERLYTGGPLTNSKQNGVRRORA 2700
QY 2701 SGVLTTSNGTILCYLVKATAACRAAKLODCTMLVNGDVLVICSAGTQEDAAALRAPTE 2760
Db 2701 SGVLTTSNGTILCYLVKATAACRAAKLODCTMLVNGDVLVICSAGTQEDAAALRAPTE 2760
QY 2761 AWTRYAPPGPPPOPEVDLELITSCSNVSVAHDASGRKVVYITDPTTPLARAAWETAR 2820
Db 2761 AWTRYAPPGPPPPPEVDLELITSCSNVSVAHDASGRKVVYITDPTTPLARAAWETAR 2820
QY 2821 HTPINSWLGNTIMVAPTILWARMILMTHFFSILLAEQLEKALDCQIYACYSIEPLDLPQ 2880
Db 2821 HTPVNSWLGNTIMVAPTILWARMILMTHFFSILLAEQLEKALDCQIYACYSIEPLDLPQ 2880
QY 2881 IIEELJHLSAFTLHSYSGEINRVASCLRKUGVPLPTWRHRAVSVRKLLSQGGRAATC 2940

Db 2881 IIQRLHGLSFAHSHSVSPGHNINRASCRLKLGVPPLRVWRHRARSVRABLLSQGGRATC 2940
 QY 2941 GRYLENWAYRTCKLTPFAASQDLDSGFVAGYSGDIIYHLSLRARPRFWPLCLILLSV 3000
 Db 2941 GRYLENWAYRTCKLTPFAASQDLDSGFVAGYSGDIIYHLSLRARPRFWPLCLILLSV 3000
 QY 3001 GVGIVLLENR 3010
 Db 3001 GVGIVLLENR 3010
 RESULT 12
 ABG32459
 ID ABG32459 standard; protein; 3010 AA.
 XX
 AC ABG32459;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Hepatitis C virus Con 1 isolate polyprotein mutant #8.
 XX
 KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; muten;
 KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
 KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 2198 /note= "Wild-type Leu substituted by Ser"
 FT
 XX
 XX WO200259321-A2.
 EN
 XX
 PD 01-AUG-2002.
 XX
 XX 16-JAN-2002; 2002WO-RP000526.
 PF
 XX
 PR 23-JAN-2001; 2001US-0263479P.
 XX
 XX (RICE-) 1ST RICECHE BIOL MOLECULAR ANGELETTI.
 PA
 XX
 PI De Francesco R, Migliaccio G, Ponessa G;
 XX
 XX W21; 2002-599793/64.
 DR
 XX
 PT New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
 PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
 PT ribosome entry site (IRES) region, useful in studying HCV replication and
 PT expression.
 XX
 PS Claim 1; Page; 69pp; English.
 XX
 CC The invention relates to nucleic acid molecules comprising altered HCV
 CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
 CC internal ribosome entry site (IRES) region coding for one or more NS3,
 CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
 CC are detailed in the specification. Also included are (1) an expression
 CC vector comprising a nucleotide sequence coding for the altered nucleic
 CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
 CC recombinant cell human hepatoma cell comprising the altered nucleic acids
 CC ; (3) a recombinant cell produced by introducing into a human hepatoma
 CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
 CC replicon enhanced cell or which containing a functional HCV replicon; (5)
 CC an HCV replicon enhanced cells made in the method; and (6) measuring the
 CC ability of a compound to affect HCV activity. The HCV replicons and HCV
 CC replicon enhanced cells are useful in studying HCV replication and
 CC expression, and HCV and host cell interactions, producing HCV RNA and
 CC proteins, and providing a system for measuring the ability of a compound
 CC to modulate one or more HCV activities e.g. to discover drugs which may
 CC treat HCV mediated diseases such as liver failure, cirrhosis and
 CC hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
 CC polyprotein (comprising the Core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A

CC and NS5B proteins), NS5A mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the HCV sequence appearing as ABG32451 and the information in claim
 CC 1
 XX
 SQ Sequence 3010 AA;
 Query Match 95.7%; Score 15317; DB 5; Length 3010;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 2851; Conservative 76; Mismatches 83; Indels 0; Gaps 0;
 QY 1 MSTNPKPQRTKRNTRRPQDVKFPGGQIVGGVYLLPRRGRJGVRAIRKASRSQPRG 60
 Db 1 MSTNPKPQRTKRNTRRPQDVKFPGGQIVGGVYLLPRRGRJGVRAIRKASRSQPRG 60
 QY 61 R3QPIPKARRPEGRAWAQPGYWPPLYGNEGLGWAGWLLSPRGSPSWGPTDPRRSRNLG 120
 Db 61 R3QPIPKARRPEGRAWAQPGYWPPLYGNEGLGWAGWLLSPRGSPSWGPTDPRRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYTPLVGAPLGGARALAHGVRLVEDGVNATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYTPLVGAPLGGARALAHGVRLVEDGVNATGNLPGCSFSIFLLA 180
 QY 181 LLSCLITIPASAYEVRNVSGIYHVVTNDCSNSSIVVEAADVIMETPCVPCVQEGNSSRCWV 240
 Db 181 LLSCLITIPASAYEVRNVSGIYHVVTNDCSNSSIVVEAADVIMETPCVPCVQEGNSSRCWV 240
 QY 241 ALTPTLAARNASVPTTIRRHVLDLVGTAAFCPSAMYVDLGGSIPLYSQLTPTPRRHET 300
 Db 241 ALTPTLAARNASVPTTIRRHVLDLVGTAAFCPSAMYVDLGGSIPLYSQLTPTPRRHET 300
 QY 301 VQDCNCSIYPGHVSGHRMAMDMNWSPTTALVVSQLLRIPQAVVDVAVAGAHWGLAGLA 360
 Db 301 VQDCNCSIYPGHVSGHRMAMDMNWSPTTALVVSQLLRIPQAVVDVAVAGAHWGLAGLA 360
 QY 361 YYSVGNWAKVLIIVALLPAGVDGTHHTTRVAGHTTSGFTSLPSSGASQKIOLVNTNGSW 420
 Db 361 YYSVGNWAKVLIIVALLPAGVDGTHHTTRVAGHTTSGFTSLPSSGASQKIOLVNTNGSW 420
 QY 421 HINRTALNCNDSLOTGFFAALFYAHKENSNGCPCPMASCRPIDWFAQKGPITTKENSS 480
 Db 421 HINRTALNCNDSLOTGFFAALFYAHKENSNGCPCPMASCRPIDWFAQKGPITTKENSS 480
 QY 481 DQRYCWHYAPRCPGCVVPASQVGPVYCFTPSPVWVGTDRSGVPTYSWGENEDVMLN 540
 Db 481 DQRYCWHYAPRCPGCVVPASQVGPVYCFTPSPVWVGTDRSGVPTYSWGENEDVMLN 540
 QY 541 NTRPQGNWFGCTWNNSTGFTKTCGGPPONIGGVGNNTLICPTDCFRKHPEATYTKCGSG 600
 Db 541 NTRPQGNWFGCTWNNSTGFTKTCGGPPONIGGVGNNTLICPTDCFRKHPEATYTKCGSG 600
 QY 601 PWLTPRCCLVDYPYRLWHYPTLNFPIFKVRMYGVGVHRLNAACNTRGRCNLEDDRS 660
 Db 601 PWLTPRCCLVDYPYRLWHYPTLNFPIFKVRMYGVGVHRLNAACNTRGRCNLEDDRS 660
 QY 661 ELSPLLSTTTEWQILPCAPTLCPALSTGLIHLHQNIVDVOYLXGVGSFAVSPFAIKWEYIL 720
 Db 661 ELSPLLSTTTEWQILPCAPTLCPALSTGLIHLHQNIVDVOYLXGVGSFAVSPFAIKWEYIL 720
 QY 721 LLELLADARVCACLWMLLIAQEAALENLVNNAASVAGAHGILSLFVFFCAAWYIKG 780
 Db 721 LLELLADARVCACLWMLLIAQEAALENLVNNAASVAGAHGILSLFVFFCAAWYIKG 780
 QY 781 RLAPGAAYAVGVWVPHILLILLALPPRAYALDRMAASCGGAVLGVFLTLSPYKVELT 840
 Db 781 RLAPGAAYAVGVWVPHILLILLALPPRAYALDRMAASCGGAVLGVFLTLSPYKVELT 840
 QY 841 RLIIWLQYFITRAEAHMQVWVPLNVGRGDAIILLTCAVHPBELIFDITKLLAILGLPLM 900
 Db 841 RLIIWLQYFITRAEAHMQVWVPLNVGRGDAIILLTCAVHPBELIFDITKLLAILGLPLM 900
 QY 901 VLQAGITRVYFVRAQGLIRACMLVRKAGCHTVQVWVFKLGAITGVYVNHHTPDRDWA 960

Db 901 VLAQAGITKVPYFYFAEGLIRACMLVRKAVGCHYVQMALMKLAALATGTIVYDHLTPLRDWA 960
QY 961 HAGLRDLAVAVEVPFSAMETKVIETWADTAACGDIILGLPVSARGRKRIFLGPADSLEG 1020
Db 961 HAGLRDLAVAVEVPFSAMETKVIETWADTAACGDIILGLPVSARGRKRIHLGPADSLEG 1020
QY 1021 QGWRLLAPITAYSQQTRGVLGCCIITSLTGDRKNQVEGEVQVSTATQSFATCINGVCWT 1080
Db 1021 QGWRLLAPITAYSQQTRGVLGCCIITSLTGDRKNQVEGEVQVSTATQSFATCINGVCWT 1080
QY 1081 VHGAGSKTLAGKPIITOMYTNVDLDLVGQWAPPGARGMTPCSCSSDLYLVTRADVI 1140
Db 1081 VHGAGSKTLAGKPIITOMYTNVDLDLVGQWAPPGARGMTPCSCSSDLYLVTRADVI 1140
QY 1141 PVRRRGDSRGLSPRPVSYLKGSSGPLLCPSGHVGVFRAAVCTRGVAKAVDFIPVES 1200
Db 1141 PVRRRGDSRGLSPRPVSYLKGSSGPLLCPSGHAVGIPRAVCTRGVAKAVDFIPVES 1200
QY 1201 METTMRSPVFTDNSTPAPVQTPQVAHLHAPTQSGKSTKVPAAAYAGKYVLVLPNSVAA 1260
Db 1201 METTMRSPVFTDNSTPAPVQTPQVAHLHAPTQSGKSTKVPAAAYAGKYVLVLPNSVAA 1260
QY 1261 TLGFGAYMSKAGIGDENITGVRTITTGSGITVSTYKELADGCGSGGAYDIIICECHS 1320
Db 1261 TLGFGAYMSKAGIGDENITGVRTITTGSGITVSTYKELADGCGSGGAYDIIICECHS 1320
QY 1321 TDSTTILGIGTVLDOAETAGARLVLATATPPGSVTVPHPNIBIEIGLSNKGRIPIFYGKAI 1380
Db 1321 TDSTTILGIGTVLDOAETAGARLVLATATPPGSVTVPHPNIBIEIGLSNKGRIPIFYGKAI 1380
QY 1381 PLEAKGGRHLFCHSKKXKDELAALKLGLGNNAVAYRGLDVSVIPTGDDVVVWATDAL 1440
Db 1381 PLETTKGRHLFCHSKKXKDELAALKLGLGNNAVAYRGLDVSVIPTGDDVVVWATDAL 1440
QY 1441 MTGFTGDFSDVDCNTCVTQVDFSDPTFTTETTTVPQDAVSRSORRGRTGCRSGIYR 1500
Db 1441 MTGFTGDFSDVDCNTCVTQVDFSDPTFTTETTTVPQDAVSRSORRGRTGCRSGIYR 1500
QY 1501 FVTGGERPMSGFSSVLCYCDAGCAWYELTPAETSRLRAYLNTPLGVPQODHLEFVES 1560
Db 1501 FVTGGERPMSGFSSVLCYCDAGCAWYELTPAETSRLRAYLNTPLGVPQODHLEFVES 1560
QY 1561 VFTGLTHIDAHFLSQKQAGDNPPYLVAQATVCAQAQAPPSWDQMWKCLRKLKPTLHG 1620
Db 1561 VFTGLTHIDAHFLSQKQAGDNPPYLVAQATVCAQAQAPPSWDQMWKCLRKLKPTLHG 1620
QY 1621 PTPLLYRLGAVQNEVLTTHPIKYINWACMSADLEVVVTSTWLVGGVLAALAYCLITGVS 1680
Db 1621 PTPLLYRLGAVQNEVLTTHPIKYINWACMSADLEVVVTSTWLVGGVLAALAYCLITGVS 1680
QY 1681 VIVGRIILSGKPAIPDREVLYOEFDMEBECASQLPYIBQGMQLABQFKQKALGLLQAT 1740
Db 1681 VIVGRIILSGKPAIPDREVLYOEFDMEBECASQLPYIBQGMQLABQFKQKALGLLQAT 1740
QY 1741 KOAEAAAPVVEKRWALETFWAKHMWNFTSGIQLAGLSTLPGNPAIASIMAFASITSP 1800
Db 1741 KOAEAAAPVVEKRWALETFWAKHMWNFTSGIQLAGLSTLPGNPAIASIMAFASITSP 1800
QY 1801 LITQNTLLNIIIGGWAAQLAPPSAASAFVAGIAGAAVGSIGLGVLDVILLAGYGAGYA 1860
Db 1801 LITQNTLLNIIIGGWAAQLAPPSAASAFVAGIAGAAVGSIGLGVLDVILLAGYGAGYA 1860
QY 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHYGPEGAVQWMNRLLI 1920
Db 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHYGPEGAVQWMNRLLI 1920
QY 1921 AFASRGNHVSPTHYVPESDAAARVTOILSSLTITQLKLHWINEDCSTPCSGWLRDV 1980
Db 1921 AFASRGNHVSPTHYVPESDAAARVTOILSSLTITQLKLHWINEDCSTPCSGWLRDV 1980
QY 1981 WDWCITVLTFDKTWLQSKLLPLPVPFPLSCORGKYGWVRGDMOTTCPCGAQITAGHYK 2040
Db 1981 WDWCITVLTFDKTWLQSKLLPLPVPFPLSCORGKYGWVRGDMOTTCPCGAQITAGHYK 2040

RESULT 13

ABC32455

ID ABG32455 standard; protein; 3010 AA.

QY 2041 NGSXRIVGERTCSNTWEGTFPINAYTTGCTPSPAPNYSRALWRVAABEYVEVTRVGDSH 2100
Db 2041 NGSXRIVGERTCSNTWEGTFPINAYTTGCTPSPAPNYSRALWRVAABEYVEVTRVGDSH 2100
QY 2101 YVTGMTDNNVKCPQVPAPEFFTEVDGVRILHEYAPACKPILLREDVTFOVLGNQYLVGSQL 2160
Db 2101 YVTGMTDNNVKCPQVPAPEFFTEVDGVRILHEYAPACKPILLREDVTFOVLGNQYLVGSQL 2160
QY 2161 PCEPEPDVTLTSMLTDSHITAEAKRLARCSPPSLASSASQI SAPSLKATCTTHHD 2220
Db 2161 PCEPEPDVAVLTSMLTDSHITAEAKRLARCSPPSLASSASQI SAPSLKATCTTHHD 2220
QY 2221 SPDADLIEANLWRQBMGNGNITRVESENKVILDSFEPLHABGDREISVAEAILKRSK 2280
Db 2221 SPDADLIEANLWRQBMGNGNITRVESENKVILDSFEPLHABGDREISVAEAILKRSK 2280
QY 2281 PFSALPIWARPDYNNPPLLESWKDPDYPVPVHVGCELPPTKAPPIPPPRKRTVVLTESNV 2340
Db 2281 PFRAMPWARPDYNNPPLLESWKDPDYPVPVHVGCELPPTKAPPIPPPRKRTVVLSESTV 2340
QY 2341 SSALAEIATKTFGSSGSSAVDSGTATAPDLASDDGDKGSDVESYSSMPPLEGECPDPL 2400
Db 2341 SSALAEIATKTFGSSGSSAVDSGTATAPDLASDDGDKGSDVESYSSMPPLEGECPDPL 2400
QY 2401 SDGSSWTVSEEAESDVCSSMSYTWGTALITPCAABESKLPINPLNSLLRHHNNVYAT 2460
Db 2401 SDGSSWTVSEEAESDVCSSMSYTWGTALITPCAABEYKLPINALNSLLRHHNNVYAT 2460
QY 2461 SRSASLRQKVTFFDLQVLDHRYDLVKEMKAKASTVKAKLSIIEEACKLTPPHSARSKF 2520
Db 2461 SRSASLRQKVTFFDLQVLDHRYDLVKEMKAKASTVKAKLSIIEEACKLTPPHSARSKF 2520
QY 2521 GYGAKDVNLSKAVNHRISYVWEDILETETETETETETETETETETETETETETETET 2580
Db 2521 GYGAKDVNLSKAVNHRISYVWEDILETETETETETETETETETETETETETETET 2580
QY 2581 FPDGLGVRCERKALYDVVSTLPOAVMGSSYGFQYSPKORVFLVNTWKSXKCPMGFSYDT 2640
Db 2581 FPDGLGVRCERKALYDVVSTLPOAVMGSSYGFQYSPGQVDFLVNAWAKKCPMGFAYDT 2640
QY 2641 RCFDSTVTESDIRVESIYQCCDLAPRARQALRSUTERLYIGGPLTNSKGQNCYRRCRA 2700
Db 2641 RCFDSTVTENDIRVESIYQCCDLAPRARQALRSUTERLYIGGPLTNSKGQNCYRRCRA 2700
QY 2701 SGVLTITSCNTLITCVLKATAACRAAKLODCTMLVNGDDLVVICESAGTOEAAAARAFTE 2760
Db 2701 SGVLTITSCNTLITCVLKATAACRAAKLODCTMLVNGDDLVVICESAGTOEAAAARAFTE 2760
QY 2761 AMTRYSAFPDPPPEYDLELITSCSSNVSAHDASGRVYVYLTDRPTTPLARAAWETAR 2820
Db 2761 AMTRYSAFPDPPPEYDLELITSCSSNVSAHDASGRVYVYLTDRPTTPLARAAWETAR 2820
QY 2821 HTPINSWLGNIIMYAPTWARMLMTHFPFSLIAQOELEKALDCOIYACYSIEPLDLPQ 2880
Db 2821 HTPVNSWLGNIIMYAPTWARMLMTHFPFSLIAQOELEKALDCOIYACYSIEPLDLPQ 2880
QY 2881 IIERLHGLSAFTLHSGYSGEINVASCLRKLVPLRTWRHRARSVRKALLSQGGRAATC 2940
Db 2881 IIOQLHGLSATSLSHSGEINVASCLRKLVPLRTWRHRARSVRKALLSQGGRAATC 2940
QY 2941 GRYLENVAVRTKLLKLTPIPAASQDLSCGFVAGYSGGDIYHSLSRARPRWPPLCLLLLSV 3000
Db 2941 GYLENVAVRTKLLKLTPIPAASQDLSCGFVAGYSGGDIYHSLSRARPRWPPLCLLLLSV 3000
QY 3001 GVGIYLLPNR 3010
Db 3001 GVGIYLLPNR 3010

XX AC ABC32455;
 XX DT 15-NOV-2002 (first entry)
 XX DE Hepatitis C virus Con 1 isolate polyprotein mutant #4.
 XX KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; mutain; HCV;
 XX KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
 XX KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
 XX OS Hepatitis C virus.
 XX OS Synthetic.
 XX Key Location/Qualifiers
 XX FH Misc-difference 2041
 XX FT /note= "Wild-type Asn substituted by Thr"
 XX PN WO200259321-A2.
 XX PD 01-AUG-2002.
 XX PF 16-JAN-2002; 2002WO-BF000526.
 XX PR 23-JAN-2001; 2001US-0263479P.
 XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
 XX PA De Francesco R, Migliaccio G, Paonessa G;
 XX WPI; 2002-599793/64.
 XX New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
 XX NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
 XX ribosome entry site (IRES) region, useful in studying HCV replication and
 XX expression.
 XX Claim 1; Page; 69pp; English.
 XX The invention relates to nucleic acid molecules comprising altered HCV
 XX NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
 XX internal ribosome entry site (IRES) region coding for one or more NS3,
 XX NS5A, or EMCV IRES mutations, respectively. The location of the mutations
 XX are detailed in the specification. Also included are (1) an expression
 XX vector comprising a nucleotide sequence coding for the altered nucleic
 XX acids, which is transcriptionally coupled to an exogenous promoter; (2) a
 XX recombinant cell human hepatoma cell comprising the altered nucleic acids
 XX ; (3) a recombinant cell produced by introducing into a human hepatoma
 XX cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
 XX replicon enhanced cell or which containing a functional HCV replicon; (5)
 XX an HCV replicon enhanced cells made in the method; and (6) measuring the
 XX ability of a compound to affect HCV activity. The HCV replicons and HCV
 XX replicon enhanced cells are useful in studying HCV replication and
 XX expression, and HCV and host cell interactions, producing HCV RNA and
 XX proteins, and providing a system for measuring the ability of a compound
 XX to modulate one or more HCV activities e.g. to discover drugs which may
 XX treat HCV mediated diseases such as liver failure, cirrhosis and
 XX hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
 XX polyprotein (comprising the Core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A
 XX and NS5B proteins), NS5A mutant of the invention. Note: The present
 XX sequence is not shown in the specification but was created by the indexer
 XX using the HCV sequence appearing as ABC32451 and the information in claim
 XX 1
 XX Sequence 3010 AA;
 XX Query Match 95.7%; Score 15317; DB 5; Length 3010;
 XX Best Local Similarity 94.7%; Pred. No. 0;
 XX Matches 2851; Conservative 76; Mismatches 83; Indels 0; Gaps 0;
 QY 1 MSTNFKPQKTKNTNRRPDQVFKPGGQIVGGVYLLPRRGRGLGVRATRKASRSQRG 60
 DB 1 MSTNFKPQKTKNTNRRPDQVFKPGGQIVGGVYLLPRRGRGLGVRATRKTSRSQRG 60

QY 61 RRQPIKARPEGRAWAQPGYFWPLYGNEGLGWAQWLLSPGRSRPSWGPTDPRRSRNIG 120
 DB 61 RRQPIKARPEGRAWAQPGYFWPLYGNEGLGWAQWLLSPGRSRPSWGPTDPRRSRNIG 120
 QY 121 KVIDTLTCGFADLMGYIPLVAGAPLGAARALAHGVRLVEDGVVYATGNLPCSCSFIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPLVAGAPLGAARALAHGVRLVEDGVVYATGNLPCSCSFIFLLA 180
 QY 181 LLSCLTIPASAYEVNRVSGIYEVNTDCSNSSIVTEAADVIMETPCGVCVCEGNSRCWV 240
 DB 181 LLSCLTIPASAYEVNRVSGVYVHTNDCSNASIVTEAADMIMETPCGVCVCEGNSRCWV 240
 QY 241 ALTPTLAARNASVPTTIRRHVDLLLVGTAAPCSAMYGDLCSIFLVSLQTLFSPRRHET 300
 DB 241 ALTPTLAARNASVPTTIRRHVDLLLVGTAALCSAMYVGLDGLSVFLVAQLFTFSPRRHET 300
 QY 301 VQDCNCSYPGHVSCHRMADMMNWSPTTALVVSQILLRIPOAVVDVYVAGAHGWGLAGLA 360
 DB 301 VQDCNCSYPGHVTGHRMADMMNWSPTTALVVSQILLRIPOAVVDVYVAGAHGWGLAGLA 360
 QY 361 YYSWGNWAKVLIIVALLPAGVDGTHHTGRVAGHTTSGFTSLFSGSGASOKTLQVNTGWSW 420
 DB 361 YYSWGNWAKVLIIVMLLPAGVDGTYVTGTMAXNTLIGTSLFSPGSSOKTLQVNTGWSW 420
 QY 421 HINRTALNCNDSLOTPFAALFYAHKENS CGPERMASCRPIDWPAQWGPTITTKPNS 480
 DB 421 HINRTALNCNDSLNTGFLAALFYVHKFNSSCGPERMASCSPIDAPAQWGPTITTESHS 480
 QY 481 DQPYCWHYAPRPGVPEASQVCGPVYCTFSPVVVGTTRSGVPTYSWGENEDVMLLN 540
 DB 481 DQPYCWHYAPRPGVPEAAQVCGPVYCTFSPVVVGTTRSGVPTYSWGENEDVMLLN 540
 QY 541 NTRPPQGNWFCTWNSTGFTKTCGPPCNIGGVNRTLCPTDCFRKHPBATTYKCGSG 600
 DB 541 NTRPPQGNWFCTWNSTGFTKTCGPPCNIGGIGNKTLCTPTDCFRKHPBATTYKCGSG 600
 QY 601 PWLTPRCIVDYFYLWHYPCITLNFSTFKVRYMGVGVHEHRLNAACNWRGECNLEDRS 660
 DB 601 PWLTPRCIVDYFYLWHYPCITVNFSTFKVRYMGVGVHEHRLNAACNWRGECNLEDRS 660
 QY 661 ELSPLLSTTWEQILPCAFPTTLPALSTGLIHLHQNIVDVQVLYGVGSAFVSFAIKWEVIL 720
 DB 661 ELSPLLSTTWEQVLPSTTLPALSTGLIHLHQNIVDVQVLYGVGSAFVSFAIKWEVIL 720
 QY 721 LFLLLADARVACILWMLLIQAQEALENIWVNAASVAGAHGILSFLVFFCAAWYIKG 780
 DB 721 LFLLLADARVACILWMLLIQAQEALENIWVNAASVAGAHGILSFLVFFCAAWYIKG 780
 QY 781 RLAPGAAYAFYGVNPLLLIALLPRAVALDRENAASCGGAVLYGIVPLTILSPYKVELT 840
 DB 781 RLVFGAAYALYGVNPLLLIALLPRAVAMDRENAASCGGAVFVGLILLTILSPHYKFLA 840
 QY 841 RLIVWLYFITRABAHQVWVPPVPLNVRGGRDAIILLTCAVEPELIFDITKLLAILGLPM 900
 DB 841 RLIVWLYFITRABAHQVWVPPVPLNVRGGRDAVILLTCAIHPILFITITKLLAILGLPM 900
 QY 901 VLQAGITRVYFVTAQGLIRACMLVRKAGHVYQVMFKLGAITGYVYNNHTLTPBDWA 960
 DB 901 VLQAGITRVYFVTAQGLIRACMLVRKAGHVYQVMFKLGAITGYVYNNHTLTPBDWA 960
 QY 961 HAGLRDLAVALVEPVVFSAMETKVTWGADTAACGDIILGTPVARSKEFLGPAADSLG 1020
 DB 961 HAGLRDLAVALVEPVVFSAMETKVTWGADTAACGDIILGTPVARSKEFLGPAADSLG 1020
 QY 1021 QGWELLAPITAYSQOTRGVLGCIITSLTGRDKQVGEVQVWSTATOSFLATCINGCWT 1080
 DB 1021 QGWELLAPITAYSQOTRGVLGCIITSLTGRDKQVGEVQVWSTATOSFLATCINGCWT 1080
 QY 1081 VYHAGSKTLAGPKGPTQMTYVNDLVGVQAPPGARMPSCSGSSDLYLVTRHADVI 1140
 DB 1081 VYHAGSKTLAGPKGPTQMTYVNDLVGVQAPPGARMPSCSGSSDLYLVTRHADVI 1140

1141 PVRARGDSRLSPRVSYLKGSSGGPLLCPSGHVGVGFRAAAVCTRGVAKAVDFPVES 1200
1141 PVRARGDSRLSPRVSYLKGSSGGPLLCPSGHAGVGFRAAVCTRGVAKAVDFPVES 1200
1201 METMRSFVTDNSTPAVCTCFVAHLHAPTSGSKTKVPAAYAAAGYKVLVNSVAA 1260
1201 METMRSFVTDNSTPAVCTCFVAHLHAPTSGSKTKVPAAYAAAGYKVLVNSVAA 1260
1261 TLGFGAYMSKAHGDIDNIRGTITTTGGGSIITYSTYKGPLADGGCSGGAYDIIICDBCHS 1320
1261 TLGFGAYMSKAHGDIDNIRGTITTTGAPITVSTYKGPLADGGCSGGAYDIIICDBCHS 1320
1321 TDSITLIGITVLDQABTAGARLVVLTATP2GSVTVPHNPIEIGLSNNGEIPFYGKAI 1380
1321 TDSITLIGITVLDQABTAGARLVVLTATP2GSVTVPHNPIEIVALSSTGEIPFYGKAI 1380
1381 PIETIKGRLIIFCHSKKDELAALKITGLGNAAVAYRGLDVSVTPPIGDVVVATDAL 1440
1381 PIETIKGRLIIFCHSKKDELAALKUSGLGNAAVAYRGLDVSVTPSGDVIIVATDAL 1440
1441 MTGFTGDFSDVDCNTCVTVQTVDFSLDPTFTIETTVPQDAVRSQRRGRTGRGSIYR 1500
1441 MTGFTGDFSDVDCNTCVTVQTVDFSLDPTFTIETTVPQDAVRSQRRGRTGRGSIYR 1500
1501 FVTPGPRPSGMPDSSVLCEDYDAGCAWYELTPAETSVRLPAYLNTPLGLPVCQDHLFWES 1560
1501 FVTPGPRPSGMPDSSVLCEDYDAGCAWYELTPAETSVRLPAYLNTPLGLPVCQDHLFWES 1560
1561 VFTGLTHIDAHFLSQTKQAGDNPPYLVAQATVCAQAQAPPPSDMDQWKILRLKPTLHG 1620
1561 VFTGLTHIDAHFLSQTKQAGDNPPYLVAQATVCAQAQAPPPSDMDQWKILRLKPTLHG 1620
1621 PTPLLYRLGAVQNEVLTHTPIKYIMACMSADLEVTSTWLVGGVLAALAAVCLTGTGV 1680
1621 PTPLLYRLGAVQNEVLTHTPIKYIMACMSADLEVTSTWLVGGVLAALAAVCLTGTGV 1680
1681 VLVGRILILSGKPAVDPREVLYQBPPEMBECASQLPIYISQGMOLABQFKQKALGLLOTAT 1740
1681 VLVGRILILSGKPAIIPDRVLYREFDEMBECASHLYPIEQGMOLABQFKQKALGLLOTAT 1740
1741 KOAEAAAPVVEGSKRALETFWAKHMNFISGIQYLAGLSTLPCNPALIASIMAFASITSP 1800
1741 KOAEAAAPVVEGSKRTLEAFNAKHMNFISGIQYLAGLSTLPCNPALIASIMAFASITSP 1800
1801 LTTQNTLLNIILGGWVAALAPPASAASAFVAGIAGAAGVSGIGLKVLDVILLAGYGAGVA 1860
1801 LTTQNTLLNIILGGWVAALAPPASAASAFVAGIAGAAGVSGIGLKVLDVILLAGYGAGVA 1860
1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAGVQWNRLLI 1920
1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAGVQWNRLLI 1920
1921 AFASRGNHVSPTHTYVPESDAAARVQILSSLTITQLLKLHOWINEDCSTPCSGSWLRDV 1980
1921 AFASRGNHVSPTHTYVPESDAAARVQILSSLTITQLLKLHOWINEDCSTPCSGSWLRDV 1980
1981 KDWICTVLTDFXTWLOSULLPDLPGVFFLSQCRGYKGVWRGDCIMQTTCCPCAQIAGHVK 2040
1981 KDWICTVLTDFXTWLOSULLPDLPGVFFLSQCRGYKGVWRGDCIMQTTCCPCAQIAGHVK 2040
2041 NGSMRIIVGPRTCSTNTHGTFPBNAYVTGCTPSAPNYSRALWRVAABEYVETVTRVGDPH 2100
2041 TGSMRIIVGPRTCSTNTHGTFPBNAYVTGCTPSAPNYSRALWRVAABEYVETVTRVGDPH 2100
2101 YVTGMTDNVCKFCQVPAPEFFTEVDGVRHLHYA2ACKPILLREDVTFFQVGLNQYLVGSOL 2160
2101 YVTGMTDNVCKFCQVPAPEFFTEVDGVRHLHYA2ACKPILLREDVTFFQVGLNQYLVGSOL 2160
2161 PCEPPEPDVTLTSMITDPSHITAFATKRLARGSPSPSLASSASQLSAPSLKATCTTHD 2220
2161 PCEPPEPDVTLTSMITDPSHITAFATKRLARGSPSPSLASSASQLSAPSLKATCTTHD 2220
2221 SPDADLIEANLLWRQMGNIIRVSEENKVILDSFEPLHAEGDEREISVAEEILKRK 2280

1221 SPDADLIEANLLWRQMGNIIRVSEENKVILDSFEPLHAEGDEREISVABILRSRK 2280
2281 PFSALPIWARPDYNPPLBSWKDPDYPVPVHGCPPLPTKAPPITPPRRKRTVLTESNV 2340
2281 FFRAMP1WARPDYNPPLBSWKDPDYPVPVHGCPPLPAKAPPIPPRRKRTVLTSESTV 2340
2341 SSALAEALAKTKTGGSSGSAVDSGTATALPOLASDDGKSDVESISSMPLGEPGDDPL 2400
2341 SSALAEALAKTKTGGSSGSAVDSGTATASPDQSDGSDGSDVESISSMPLGEPGDDPL 2400
2401 SDGSWSTVEEASEDEVVCCSMSTWTGALITPCAAEESKLPINPLNSLRLHNNMVIYAT 2460
2401 SDGSWSTVEEASEDEVVCCSMSTWTGALITPCAAEETKLPINALNSLRLHNNMVIYAT 2460
2461 SRSASLRQKVTFRDLQVLDHRYDVLKEMKAKASTVKAKLLSVEEACKLTPPHSARSKF 2520
2461 SRSASLRQKVTFRDLQVLDHRYDVLKEMKAKASTVKAKLLSVEEACKLTPPHSARSKF 2520
2521 GYGAKDVRNLSSRAVNHHSVNEDELLEDDETETIDTITMAKNVFCVQPEKGGKPARLIV 2580
2521 GYGAKDVRNLSSRAVNHHSVNEDELLEDDETETIDTITMAKNVFCVQPEKGGKPARLIV 2580
2581 FPDLGVRVCEKMAIXDVVSTLPAQVWGSYGFQYSPKORVEELVNTWKSCKCPMGFSYDT 2640
2581 FPDLGVRVCEKMAIXDVVSTLPAQVWGSYGFQYSPGQVVEFLVNAWAKKCPMGFAYDT 2640
2641 RCPDSTVIESDTRVBESIYQCCDLAPARQATRSUTERLYIGGPLTNSKGQNGYRCHA 2700
2641 RCPDSTVIESDTRVBESIYQCCDLAPARQATRSUTERLYIGGPLTNSKGQNGYRCHA 2700
2701 SGVLTTSCNTLTCVLKATAACRAAKLOCTMLVNGDDOLVWICESAGTOEDAAALRAFTE 2760
2701 SGVLTTSCNTLTCVLKATAACRAAKLOCTMLVNGDDOLVWICESAGTOEDAAALRAFTE 2760
2761 AMTRYSAAPPDPPQPEYDLELITSCSNVSVVAHDASGKRVYVLTDRPTTPLARAANETAR 2820
2761 AMTRYSAAPPDPPQPEYDLELITSCSNVSVVAHDASGKRVYVLTDRPTTPLARAANETAR 2820
2821 HTPINSWLNIIIMYAPTWARMLTHFPFSLIAQOELEKALDCOYIYACYSIEPLDLFQ 2880
2821 HTPINSWLNIIIMYAPTWARMLTHFPFSLIAQOELEKALDCOYIYACYSIEPLDLFQ 2880
2881 ITERLHGLSAFTLHGSYSGEINRVASCLURKLGVPPLRTWRHRARSVRKLLSQGGRAATC 2940
2881 ITERLHGLSAFTLHGSYSGEINRVASCLURKLGVPPLRTWRHRARSVRKLLSQGGRAATC 2940
2941 GRYLNNWAVRTKLTPTIPAAASQDLDSGFVAGYSGGDIYHLSLRARPRWFPCLILLSV 3000
2941 GRYLNNWAVRTKLTPTIPAAASQDLDSGFVAGYSGGDIYHLSLRARPRWFPCLILLSV 3000
3001 GVGIYLLNRR 3010
3001 GVGIYLLNRR 3010
RESULT 14
ABG32457
ID ABG32457 standard; protein; 3010 AA.
AC ABG32457;
XX
DT 15-NOV-2002 (first entry)
XX
DE Hepatitis C virus Con 1 isolate polyprotein mutant #6.
XX
KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; mutcin;
KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
XX internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
CS Hepatitis C virus.
XX Synthetic.

FH Key Location/Qualifiers
 FT Misc-difference 2173 /note= "Wild-type Phe substituted by Ser"
 XX
 PN W0200259321-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 16-JAN-2002; 2002WO-BP000526.
 XX
 PR 23-JAN-2001; 2001US-0263479P.
 XX
 PA (SICE-) IST RICERCHER BIOL MOLECOLARE ANGELETTI.
 XX
 PI De Francesco R, Migliaccio G, Paonessa G;
 XX
 DR WPI; 2002-599793/64.
 XX
 DR New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
 PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
 PT ribosome entry site (IRES) region, useful in studying HCV replication and
 PT expression.
 PS
 PS Claim 1; Page; 69pp; English.
 CC
 CC The invention relates to nucleic acid molecules comprising altered HCV
 CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
 CC internal ribosome entry site (IRES) region coding for one or more NS3,
 CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
 CC are detailed in the specification. Also included are (1) an expression
 CC vector comprising a nucleotide sequence coding for the altered nucleic
 CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
 CC recombinant cell human hepatoma cell comprising the altered nucleic acids
 CC ; (3) a recombinant cell produced by introducing into a human hepatoma
 CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
 CC replicon enhanced cell or which containing a functional HCV replicon; (5)
 CC an HCV replicon enhanced cells made in the method; and (6) measuring the
 CC ability of a compound to affect HCV activity. The HCV replicons and HCV
 CC replicon enhanced cells are useful in studying HCV replication and
 CC expression, and HCV and host cell interactions, producing HCV RNA and
 CC proteins, and providing a system for measuring the ability of a compound
 CC to modulate one or more HCV activities e.g. to discover drugs which may
 CC treat HCV mediated diseases such as liver failure, cirrhosis and
 CC hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
 CC polyprotein (comprising the Core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A
 CC and NS5B proteins), NS5A mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the HCV sequence appearing as ABG32451 and the information in claim
 CC 1
 XX
 XX Sequence 3010 AA;
 SQ
 Query Match 95.7%; Score 15317; DB 5; Length 3010;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 2851; Conservative 76; Mismatches 83; Indels 0; Gaps 0;
 QY 1 MSTNPKPORKXNTNRRPQVKGQIVGGVLLPRRPRIGVRAATKASRSQPRG 60
 DB 1 MSTNPKPORKXNTNRRPQVKGQIVGGVLLPRRPRIGVRAATKASRSQPRG 60
 QY 61 RRQIPKARRPEGRAWAQGYWPWLYGNEGIGWAGWLLSPRGSRPSWGPTDPRRSNRLG 120
 DB 61 RRQIPKARRPEGRAWAQGYWPWLYGNEGIGWAGWLLSPRGSRPSWGPTDPRRSNRLG 120
 QY 121 KVIDTLTCGFADLMGYIELVGAPLGGAAALAHGVRLVDGVNATGNLPGCSIFILIA 180
 DB 121 KVIDTLTCGFADLMGYIELVGAPLGGAAALAHGVRLVDGVNATGNLPGCSIFILIA 180
 QY 181 LLSCLTIPASAYERNVSGIIVHVNDCNSNITVEADNIMHTPCVPCVCGNSRRCW 240
 DB 181 LLSCLTIPASAYERNVSGIIVHVNDCNSNITVEADNIMHTPCVPCVCGNSRRCW 240
 QY 241 ALTPTLAARNASVPTTIRRHVDLLVGAALCSANYVGDLCGSFLVQLQFTFSRPHET 300

DB 241 ALTPTLAARNASVPTTIRRHVDLLVGAALCSANYVGDLCGSFLVQLQFTFSRPHET 300
 QY 301 VQDCNCSIYFGHVSCHRMAMDMNMNSPTTALVWSQLLRIPQAVVDMVAGAHGVLGAGLA 360
 DB 301 VQDCNCSIYFGHVSCHRMAMDMNMNSPTTALVWSQLLRIPQAVVDMVAGAHGVLGAGLA 360
 QY 361 YYSVGNWAKVLIIVALLFAGVDGETHTTGRVAGHTTSGFTSLFSSGASOKIOLVNTNSW 420
 DB 361 YYSVGNWAKVLIIVALLFAGVDGETHTTGRVAGHTTSGFTSLFSSGASOKIOLVNTNSW 420
 QY 421 HINRTALNCNDSLOTGPTFAALFYAHKFNSSGCPERMASCRPIDMFAQCGWPTTYPKNSS 480
 DB 421 HINRTALNCNDSLOTGPTFAALFYAHKFNSSGCPERMASCRPIDMFAQCGWPTTYPKNSS 480
 QY 481 DORPYCMHYAPRPGVVPASQVCGPVYCTFSPVVGTTDRSGVPTYSWGENEDVMLLN 540
 DB 481 DORPYCMHYAPRPGVVPASQVCGPVYCTFSPVVGTTDRSGVPTYSWGENEDVMLLN 540
 QY 541 NTRPPQGNWFGCTWNSNLTGFTKTCGGPPCNIGGVNRTLICPTDCFRKHPEATYTKCGSG 600
 DB 541 NTRPPQGNWFGCTWNSNLTGFTKTCGGPPCNIGGVNRTLICPTDCFRKHPEATYTKCGSG 600
 QY 601 PWTLPRLVDPYRLWHYPCTLNFSTFKVRMYVGVGERLNAACNWTGERCNLEDRDRS 660
 DB 601 PWTLPRLVDPYRLWHYPCTLNFSTFKVRMYVGVGERLNAACNWTGERCNLEDRDRS 660
 QY 661 ELSPLLLSTTTEWOLPCAPPTLPAISLGLIHLHQNIVDVQVLYGVGSAFVGFALKWEYIL 720
 DB 661 ELSPLLLSTTTEWOLPCAPPTLPAISLGLIHLHQNIVDVQVLYGVGSAFVGFALKWEYIL 720
 QY 721 LFLLLADARVACACLWMLLIQAQEAALENIVLNAASVAGAHGLSFLVFFCAAWIKG 780
 DB 721 LFLLLADARVACACLWMLLIQAQEAALENIVLNAASVAGAHGLSFLVFFCAAWIKG 780
 QY 781 RLAPCAAYAFYGVWPLLLLLALPRAYALDRENAASCGGAVLVGLVFLTSLSPYKYVFLT 840
 DB 781 RLAPCAAYAFYGVWPLLLLLALPRAYALDRENAASCGGAVLVGLVFLTSLSPYKYVFLT 840
 QY 841 ELIWLQVYFTRAAHQVWVPLNVRGGDDAIILLICAVHPELIFDITKLLAILGLPLM 900
 DB 841 ELIWLQVYFTRAAHQVWVPLNVRGGDDAIILLICAVHPELIFDITKLLAILGLPLM 900
 QY 901 VLQAGITRVPYFVRAQGLIRACMLVRKVAGHYVQVFMKLGALTGYVYNHILPLRDWA 960
 DB 901 VLQAGITRVPYFVRAQGLIRACMLVRKVAGHYVQVFMKLGALTGYVYNHILPLRDWA 960
 QY 961 HAGURDLAVALVEPVVFSAMETKVIITWGAATAACGDIILGLPVSARRGKEIFLOPADSLEG 1020
 DB 961 HAGURDLAVALVEPVVFSAMETKVIITWGAATAACGDIILGLPVSARRGKEIFLOPADSLEG 1020
 QY 1021 QGWRELLAPITAYSQOTRGVLGCIITSLTGRDKQVGEVQVVSATQSFATCINGVCWT 1080
 DB 1021 QGWRELLAPITAYSQOTRGVLGCIITSLTGRDKQVGEVQVVSATQSFATCINGVCWT 1080
 QY 1081 VYHAGAGSKTLAGPKGPTITQMTYVNDLVGVQAPPGARSMTPCSCSSDLIVTRHADVI 1140
 DB 1081 VYHAGAGSKTLAGPKGPTITQMTYVNDLVGVQAPPGARSMTPCSCSSDLIVTRHADVI 1140
 QY 1141 PVRRRGDSRGLSIPRVSILKSSGGLPLCPGSHVVGVFPRAAVCTRGVAKAVDFTPVES 1200
 DB 1141 PVRRRGDSRGLSIPRVSILKSSGGLPLCPGSHVVGVFPRAAVCTRGVAKAVDFTPVES 1200
 QY 1201 METTMRSEVFTDNPSPAVPQTFQVAHLHAPTCSGXSTKVPAAVAAQGVKVLVNSVAA 1260
 DB 1201 METTMRSEVFTDNPSPAVPQTFQVAHLHAPTCSGXSTKVPAAVAAQGVKVLVNSVAA 1260
 QY 1261 TLGFGATWSKAHGDIPNIRTVGRTIITGSGITVSTYTKFLADCGCGSGAYDIICDECHS 1320
 DB 1261 TLGFGATWSKAHGDIPNIRTVGRTIITGSGITVSTYTKFLADCGCGSGAYDIICDECHS 1320
 QY 1321 TDSITLIGITVLDQAEATAGARIWVLAATPPGSVTVPHNIEBIGLSNNGEIPFYGKAI 1380

Db 1321 TDSITLIGIVLQDAETAGARLVVLAATATPPGSSVTPHENIEEVALSSGTEIPFYGKAI 1380
 QY 1381 PIRAIKGRHLIPCHSKKKKDELAALKLTGLGNAYVYRGIDYSVIPPIDGVVVAATDAL 1440
 Db 1381 PIRAIKGRHLIPCHSKKKKDELAALKLTGLGNAYVYRGIDYSVIPPIDGVVVAATDAL 1440
 QY 1441 MTGFTGDFSDVIDNCTVOTQVDFSLDPTFTIETITVPODAVRSQRGRGTGRGSIYR 1500
 Db 1441 MTGFTGDFSDVIDNCTVOTQVDFSLDPTFTIETITVPODAVRSQRGRGTGRGSIYR 1500
 QY 1501 FVTPGEPSCWDFSSVLCECYDAGCANYELTPAETSURLRAYLNTPLGVPQDHLFEWES 1560
 Db 1501 FVTPGEPSCWDFSSVLCECYDAGCANYELTPAETSURLRAYLNTPLGVPQDHLFEWES 1560
 QY 1561 VFTGLTHIDAHFTLSQTKQAGDNPPYLVAQYATVCARAQAPPPSWDMQWKLIIRKPLTHG 1620
 Db 1561 VFTGLTHIDAHFTLSQTKQAGDNPPYLVAQYATVCARAQAPPPSWDMQWKLIIRKPLTHG 1620
 QY 1621 PTPLLYRLGAVQNEVILTHPTIKYIKACMGADLEWVITSTWLVGGVLAALAAAYCLITGVS 1680
 Db 1621 PTPLLYRLGAVQNEVILTHPTIKYIKACMGADLEWVITSTWLVGGVLAALAAAYCLITGVS 1680
 QY 1681 VIYGRILISGKPAVDPREVLYQEFDMEMECASQLPYIEQGMQIADQFKOKALGLQTAT 1740
 Db 1681 VIYGRILISGKPAIIPDREVLXEFDMEMECASHLPIYIEQMOLAEQFKOKAIGLLQTAT 1740
 QY 1741 KQAEAAAPVSVESKWRALLETFAWAKMWNFISGIOYLAGLSTLPGNPALASLMAFTASITSP 1800
 Db 1741 KQAEAAAPVSVESKWRILEAFWAKMWNFISGIOYLAGLSTLPGNPALASLMAFTASITSP 1800
 QY 1801 LITQNTILLNIGWVAAQIAPPSAASAFVAGIAGAAVSGISGLGKVLVDILAGYGAGVA 1860
 Db 1801 LITQNTILLNIGWVAAQIAPPSAASAFVAGIAGAAVSGISGLGKVLVDILAGYGAGVA 1860
 QY 1861 GALVAEKWMSGEVSTEDLVNLLPAILSPGALVGVVCAILRRHVGPBGCAVQWNRLLI 1920
 Db 1861 GALVAEKWMSGENPSTEDLVNLLPAILSPGALVGVVCAILRRHVGPBGCAVQWNRLLI 1920
 QY 1921 APASRGNHVSPTHYVPESDAAARVTOILSLTTITQLLKRHHOWINEDCSTPCSGMWLDV 1980
 Db 1921 APASRGNHVSPTHYVPESDAAARVTOILSLTTITQLLKRHHOWINEDCSTPCSGMWLDV 1980
 QY 1981 WDWICTVLDTFKTWLQSKLLPRLPFGVPFFSCQRYGKVGWGDGIMQITPCGQAQIAGHVK 2040
 Db 1981 WDWICTVLDTFKTWLQSKLLPRLPFGVPFFSCQRYGKVGWGDGIMQITPCGQAQIAGHVK 2040
 QY 2041 NSGMRITVGPRTCSNTWHGTFPINAYTTGCTPSPAPNYSRALWRVAABEVVETRVGDHF 2100
 Db 2041 NSGMRITVGPRTCSNTWHGTFPINAYTTGCTPSPAPNYSRALWRVAABEVVETRVGDHF 2100
 QY 2101 YVTGMTTDNVKPCQVPAPPEFFTEVDGVRILHRYPACKPLLRBDVTFOVGLNOLYGSOL 2160
 Db 2101 YVTGMTTDNVKPCQVPAPPEFFTEVDGVRILHRYPACKPLLRBDVTFOVGLNOLYGSOL 2160
 QY 2161 PCEPEPDVTLISMLTDPGHIITAKRIARLARGSPPSLASSASQISASLKAATCTTHHD 2220
 Db 2161 PCEPEPDVTLISMLTDPGHIITAKRIARLARGSPPSLASSASQISASLKAATCTTHHD 2220
 QY 2221 SPDAULIEANLLWRQBMGNITRVESNKVILDSPEPLHAGDDEBEISVAABILKRSK 2280
 Db 2221 SPDAULIEANLLWRQBMGNITRVESNKVILDSPEPLHAGDDEBEISVAABILKRSK 2280
 QY 2281 FFSALPIWARPDYNPPLLESWKDPDVPVPHGCPPLPTKAPPIPPRRKRTVILTESNV 2340
 Db 2281 FFSALPIWARPDYNPPLLESWKDPDVPVPHGCPPLPTKAPPIPPRRKRTVILTESNV 2340
 QY 2341 SSALAEATKTPGSSSSAVDSGTATLADDDGDKGSDVESYSMPPLGEPGDPDL 2400
 Db 2341 SSALAEATKTPGSSSSAVDSGTATLADDDGDKGSDVESYSMPPLGEPGDPDL 2400
 QY 2401 SDCGWSVTGSEASEDVVCCSMSTWGTGALITPCAASESKLPINFLNSILRHHNVYATT 2460
 Db 2401 SDCGWSVTGSEASEDVVCCSMSTWGTGALITPCAASESKLPINFLNSILRHHNVYATT 2460

QY 2461 SRSASIROKVTPDRLOVLDDHYRDVILKEMKAKASTVKAKLLSIEBACKLTTPHSKSKXF 2520
 Db 2461 SRSASIROKVTPDRLOVLDDHYRDVILKEMKAKASTVKAKLLSIEBACKLTTPHSKSKXF 2520
 QY 2521 GYGAKDVRNLSGRAVNHISVWEDLLEDTEPTDITTIMAKSEVFCVQPEKGGKPKPARLIV 2580
 Db 2521 GYGAKDVRNLSGRAVNHISVWEDLLEDTEPTDITTIMAKSEVFCVQPEKGGKPKPARLIV 2580
 QY 2581 FFDLGVVCEKMAIXYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTPKSKCPMGFSYPT 2640
 Db 2581 FFDLGVVCEKMAIXYDVVSTLPOAVMGSSYGFQYSPKQVFEFLVNTPKSKCPMGFSYPT 2640
 QY 2641 RCFDSTVTESDIRVEBSIYCCCDLAPARQAIRSLTERLYIGGPLTNSKGKCGYRCRA 2700
 Db 2641 RCFDSTVTESDIRVEBSIYCCCDLAPARQAIRSLTERLYIGGPLTNSKGKCGYRCRA 2700
 QY 2701 SGYLTTSCGNTLTCYLKATAACRAAKLQDCTMLVNGDDLVNICESAGTQEDAAALRAFTE 2760
 Db 2701 SGYLTTSCGNTLTCYLKATAACRAAKLQDCTMLVNGDDLVNICESAGTQEDAAALRAFTE 2760
 QY 2761 AMTRISAPPGDPPQBYDLELITSSCSNVSVVAHDASGKRVYVLTDRPTTDLARAAWETAR 2820
 Db 2761 AMTRISAPPGDPPQBYDLELITSSCSNVSVVAHDASGKRVYVLTDRPTTDLARAAWETAR 2820
 QY 2821 HTPINSLGNIIMYAPTILWARMILMTHPESILLAOEUKALDCCIYGACYSIEPLDLPQ 2880
 Db 2821 HTPINSLGNIIMYAPTILWARMILMTHPESILLAOEUKALDCCIYGACYSIEPLDLPQ 2880
 QY 2881 IIERLHLSAFTLHSGYSPGEINRVASCLRKLGVPPLTWRHRSVRKALLSQGGRAATC 2940
 Db 2881 IIERLHLSAFTLHSGYSPGEINRVASCLRKLGVPPLTWRHRSVRKALLSQGGRAATC 2940
 QY 2941 GRYLENWAVRTKLTPIPAASQOLDLSGFWAGYGGDIYHLSRARPFWPLCLLLLSV 3000
 Db 2941 GRYLENWAVRTKLTPIPAASQOLDLSGFWAGYGGDIYHLSRARPFWPLCLLLLSV 3000
 QY 3001 GVGIVLLPNR 3010
 Db 3001 GVGIVLLPNR 3010
 RESULT 15
 ABG32452
 ID ABG32452 standard; protein; 3010 AA.
 XX ABG32452;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Hepatitis C virus Con 1 isolate polypeptide mutant #1.
 XX
 KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; muten;
 KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
 KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 Key Location/Qualifiers
 FT Misc-difference 1095 /note= "Wild-type Gly substituted by Ala"
 FT
 XX WO200259321-A2.
 XX
 XX 01-AUG-2002.
 XX
 XX 16-JAN-2002; 2002WO-EPO00526.
 XX
 XX 23-JAN-2001; 2001US-0263479P.
 XX
 PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
 XX

De Francesco R, Migliaccio G, Paonessa G;
WPI; 2002-599793/64.
New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
ribosome entry site (IRES) region, useful in studying HCV replication and
expression.
Claim 1; Page; 69pp; English.
The invention relates to nucleic acid molecules comprising altered HCV
NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
internal ribosome entry site (IRES) region coding for one or more NS3,
NS5A, or EMCV IRES mutations, respectively. The location of the mutations
are detailed in the specification. Also included are (1) an expression
vector comprising a nucleotide sequence coding for the altered nucleic
acids, which is transcriptionally coupled to an exogenous promoter; (2) a
recombinant cell human hepatoma cell comprising the altered nucleic acids
cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
an HCV replicon enhanced cell or which containing a functional HCV replicon; (5)
ability of a compound to affect HCV activity. The HCV replicons and HCV
replicon enhanced cells are useful in studying HCV replication and
expression, and HCV and host cell interactions, producing HCV RNA and
proteins, and providing a system for measuring the ability of a compound
to modulate one or more HCV activities e.g. to discover drugs which may
treat HCV mediated diseases such as liver failure, cirrhosis and
hepatocellular carcinoma. The present sequence is the HCV replicon Cor 1
and NS5B proteins), NS3 mutant of the invention. Note: The present
sequence is not shown in the specification but was created by the indexer
using the HCV sequence appearing as ABC32451 and the information in claim
1
Sequence 3010 AA;
Query Match 95.7%; Score 15317; DB 5; Length 3010;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 2851; Conservative 76; Mismatches 83; Indels 0; Gaps 0;
QY 1 MSTNXPQKTKNTNRPODVKPPGGQIVGGVYLLPRGRLGVRAKASERQPRG 60
Db 1 MSTNXPQKTKNTNRPODVKPPGGQIVGGVYLLPRGRLGVRAKASERQPRG 60
QY 61 RRQPIPKARPEGEAWAQQPYPPLVYNEGLWAGWLLSPRGRPSWGPDPDRRERNLG 120
Db 61 RRQPIPKARPEGEAWAQQPYPPLVYNEGLWAGWLLSPRGRPSWGPDPDRRERNLG 120
QY 121 KVDTITCGFADLMGVIPVAGPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIFLLA 180
Db 121 KVDTITCGFADLMGVIPVAGPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIFLLA 180
QY 181 LLSCLTIPASAYVRNVSGHYHTNDCSNSSIVYEADVIMHTPGCPVCGENSSRCWV 240
Db 181 LLSCLTIPASAYVRNVSGHYHTNDCSNSSIVYEADVIMHTPGCPVCGENSSRCWV 240
QY 241 ALTPTLAARNASVPTTIRRHVDLLVGTAAFCAMTVGLDGSIFLVLSOLFSPRRHET 300
Db 241 ALTPTLAARNASVPTTIRRHVDLLVGTAAFCAMTVGLDGSIFLVLSOLFSPRRHET 300
QY 301 VQDCNCSIVPGHVSGRHAWMMNNSPTALVVSOLLRIPOAVVDMVAGAHGWVLAGLA 360
Db 301 VQDCNCSIVPGHVSGRHAWMMNNSPTALVVSOLLRIPOAVVDMVAGAHGWVLAGLA 360
QY 361 YYSWGNWAKVLLVALLFAGVDCGTHETTCRVAGHTTSGTSLPSSCASOKIQLVNTNGSW 420
Db 361 YYSWGNWAKVLLVALLFAGVDCGTHETTCRVAGHTTSGTSLPSSCASOKIQLVNTNGSW 420
QY 421 HINRTALNCNDSIQTOFFAALFYAHKFNSSGCPERMASCRPIDWPAQGGPITYTKENSS 480
Db 421 HINRTALNCNDSIQTOFFAALFYAHKFNSSGCPERMASCRPIDWPAQGGPITYTKENSS 480

QY 481 DORPYCWHYAPRCGVYASOVCGPYCFTPPVWVITDRSOVPTYSWGENETDMLLN 540
Db 481 DORPYCWHYAPRCGVYASOVCGPYCFTPPVWVITDRSOVPTYSWGENETDMLLN 540
QY 541 NTRPPQGNWFGCTWNNSTGFTTCGGPPCNIGGVGNRTLICPTDCFRKIPEATYTKCSG 600
Db 541 NTRPPQGNWFGCTWNNSTGFTTCGGPPCNIGGVGNRTLICPTDCFRKIPEATYTKCSG 600
QY 601 PMLTBRCLVDYPIRLWHYECTLNFSIFKVRMTVGGVEHRLNAACNWTGERCNLEDRDS 660
Db 601 PMLTBRCLVDYPIRLWHYECTLNFSIFKVRMTVGGVEHRLNAACNWTGERCNLEDRDS 660
QY 661 ELSPILLSTTEWQLPCATTTLPALSTGTHLFONIVDVOYLGVGSARVSFAIKWEYIL 720
Db 661 ELSPILLSTTEWQLPCATTTLPALSTGTHLFONIVDVOYLGVGSARVSFAIKWEYIL 720
QY 721 LLLFLLADARVACALWMLLIIAQAEAALENVLNAASVAGAHGILSLVFFCAAWYIKG 780
Db 721 LLLFLLADARVACALWMLLIIAQAEAALENVLNAASVAGAHGILSLVFFCAAWYIKG 780
QY 781 RLAPGAAYAFYGVWPLLLLLLPAYALDREMAASCGGAVLVGLVTLTSPYKVFUT 840
Db 781 RLAPGAAYAFYGVWPLLLLLLPAYALDREMAASCGGAVLVGLVTLTSPYKVFUT 840
QY 841 RLITWLOYEITRAEAHMVWVPLNVRGGRDAIILLTCAVHPBLIFDITKLLAILGLM 900
Db 841 RLITWLOYEITRAEAHMVWVPLNVRGGRDAIILLTCAVHPBLIFDITKLLAILGLM 900
QY 901 VLQAGTRVPYFVRAOGLIRACMLVKRVAGHYVQVMFKLGALGTYYVNHLPDLWA 960
Db 901 VLQAGTRVPYFVRAOGLIRACMLVKRVAGHYVQVMFKLGALGTYYVNHLPDLWA 960
QY 961 HAGLRDLAVAVBPVPSAMETKVIITWADTACGDIILGLPVSARRGKEIFIGPADSLEG 1020
Db 961 HAGLRDLAVAVBPVPSAMETKVIITWADTACGDIILGLPVSARRGKEIFIGPADSLEG 1020
QY 1021 CQWRLLAPITAYSQOTRGVLGCIITSLTGRDRKNQVEGQVGVSTATOSPLACINGVCWT 1080
Db 1021 CQWRLLAPITAYSQOTRGVLGCIITSLTGRDRKNQVEGQVGVSTATOSPLACINGVCWT 1080
QY 1081 VTHGAGSKTLAGKPGITQMYTNVDLDLVGWOAPPQASMTPCSCGSSDLVLTTHADVI 1140
Db 1081 VTHGAGSKTLAGKPGITQMYTNVDLDLVGWOAPPQASMTPCSCGSSDLVLTTHADVI 1140
QY 1141 PVRRGDSRGSLLSRPVSYLKSGSGPLLCPSGHVGVVFRAAVCTRGVAKAVDIPVES 1200
Db 1141 PVRRGDSRGSLLSRPVSYLKSGSGPLLCPSGHVGVVFRAAVCTRGVAKAVDIPVES 1200
QY 1201 METTMRSPVFTDNSTPRAVPOTFOVAHLHAPTGGSKTKVPAAYAAQGVKVLVLPNSVAA 1260
Db 1201 METTMRSPVFTDNSTPRAVPOTFOVAHLHAPTGGSKTKVPAAYAAQGVKVLVLPNSVAA 1260
QY 1261 TLGPGAYMSKAGIDPNIRTVRTITGSIITYSTYKFLADGGCGGAYDIIICDECHS 1320
Db 1261 TLGPGAYMSKAGIDPNIRTVRTITGSIITYSTYKFLADGGCGGAYDIIICDECHS 1320
QY 1321 TDSITLIGTVDQAEATAGARLVATATPPGSVTVVPHNTEELGNSGEIIPYGGAI 1380
Db 1321 TDSITLIGTVDQAEATAGARLVATATPPGSVTVVPHNTEELGNSGEIIPYGGAI 1380
QY 1381 PIETIKGRHILFCHSKKKKDELAALTKLGLNVAAYVYRGLDVSVIPPIGVWVATDAL 1440
Db 1381 PIETIKGRHILFCHSKKKKDELAALTKLGLNVAAYVYRGLDVSVIPPIGVWVATDAL 1440
QY 1441 MTGFTGDSIDSCNTCVTQTVDFSLDFTFTTETTVQDAVSRQRRGRTGRSGIYR 1500
Db 1441 MTGFTGDSIDSCNTCVTQTVDFSLDFTFTTETTVQDAVSRQRRGRTGRSGIYR 1500
QY 1501 FVTPGERPSGMESSVLCEYDAGCAWYELT2AETSIPRLAYLNTPGIPVQODHLEPWS 1560
Db 1501 FVTPGERPSGMESSVLCEYDAGCAWYELT2AETSIPRLAYLNTPGIPVQODHLEPWS 1560

QY 1561 VFTGLTHIDAHFQSOTKQAGNFPYLVAQATVCARAQAPPSQDQWKKLIRKPTLHG 1620
Db 1561 VFTGLTHIDAHFQSOTKQAGNFPYLVAQATVCARAQAPPSQDQWKKLIRKPTLHG 1620
QY 1621 PTLPLLR-GAGVNEVLHPHTTKYIMACMSADLEVVTTWLVGGVLAALAAAYCLTTGSV 1680
Db 1621 PTLPLLR-GAGVNEVLHPHTTKYIMACMSADLEVVTTWLVGGVLAALAAAYCLTTGSV 1680
QY 1681 VIVGRIILSGRPVAVDPDEVLVQSFDEMEECASOLPYIEQGMQALAEQFKQALGLQAT 1740
Db 1681 VIVGRIILSGRPVAVDPDEVLVQSFDEMEECASOLPYIEQGMQALAEQFKQALGLQAT 1740
QY 1741 KQABAAAPVSVKRALETFTWAKHMWNPISGIQVLAGLSTLPGNPAIASLMAFTASITSP 1800
Db 1741 KQABAAAPVSVKRALETFTWAKHMWNPISGIQVLAGLSTLPGNPAIASLMAFTASITSP 1800
QY 1801 LTTQNTLLFNILGGWAAQALAPPSAASAFVAGTAGAAGVSGIGLKVLDILAGYGAGVA 1860
Db 1801 LTTQNTLLFNILGGWAAQALAPPSAASAFVAGTAGAAGVSGIGLKVLDILAGYGAGVA 1860
QY 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAATILRRHVGPGECAVOMNRLI 1920
Db 1861 GALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCAATILRRHVGPGECAVOMNRLI 1920
QY 1921 AFASRGNHVSPHYVPSDAAARVTOILSLTITQLLKRLHQWNEDECSTPCSGSWLRDV 1980
Db 1921 AFASRGNHVSPHYVPSDAAARVTOILSLTITQLLKRLHQWNEDECSTPCSGSWLRDV 1980
QY 1981 WDMICTVLITDFKWLQSKLPLRGVPPFJSCQRYKGVWGRGDGIMQTCPCGQIAGHYK 2040
Db 1981 WDMICTVLITDFKWLQSKLPLRGVPPFJSCQRYKGVWGRGDGIMQTCPCGQIAGHYK 2040
QY 2041 NGSMRIVGPTTCNTWHTGTFPINAYTTGCTPPAPNYSALMRVAEEVVEVTRVGDHF 2100
Db 2041 NGSMRIVGPTTCNTWHTGTFPINAYTTGCTPPAPNYSALMRVAEEVVEVTRVGDHF 2100
QY 2101 YVTGMTDVKVPCQVPAPEFFTEVDGVRUHRVAPACKPLLRBDVTFQVGLNOYLVSQ 2160
Db 2101 YVTGMTDVKVPCQVPAPEFFTEVDGVRUHRVAPACKPLLRBDVTFQVGLNOYLVSQ 2160
QY 2161 PCEPEPDVTLTSMJLDPDPSHITAEAKRLIARGSPSLASSASQISAPLSKATCTTHD 2220
Db 2161 PCEPEPDVTLTSMJLDPDPSHITAEAKRLIARGSPSLASSASQISAPLSKATCTTHD 2220
QY 2221 SPDADLIEANLWRQEMGNI TRVESENKVILDSFEPLHAEGDEREISVAABILRSRK 2280
Db 2221 SPDADLIEANLWRQEMGNI TRVESENKVILDSFEPLHAEGDEREISVAABILRSRK 2280
QY 2281 FPSALPIWARPDYNPPLLESWKDPDYVPPVHGCPDPTKAPPIPPRRKRTVVLTESNV 2340
Db 2281 FPSALPIWARPDYNPPLLESWKDPDYVPPVHGCPDPTKAPPIPPRRKRTVVLTESNV 2340
QY 2341 SSALAEATKTFGSSSESSAVDGTATAPDQSDGDDAGSDVESYSSMPPLGERGDDPL 2400
Db 2341 SSALAEATKTFGSSSESSAVDGTATAPDQSDGDDAGSDVESYSSMPPLGERGDDPL 2400
QY 2401 SDGSMSTVSEBESAVDVCMSYTWTCALITPCAABESKLPINPISNLSLRHNNVYATT 2460
Db 2401 SDGSMSTVSEBESAVDVCMSYTWTCALITPCAABESKLPINPISNLSLRHNNVYATT 2460
QY 2461 SSSASLROKKTVDRLQVLDHVRDVLKEMKAKASTVKAALLSVBEACKLTPPHSARSKF 2520
Db 2461 SSSASLROKKTVDRLQVLDHVRDVLKEMKAKASTVKAALLSVBEACKLTPPHSARSKF 2520
QY 2521 GYGAKDVNLSRANVHRSVWEDLIEDTETPIDTTIMAKSEVFCVQPEKGGKPARLIV 2580
Db 2521 GYGAKDVNLSRANVHRSVWEDLIEDTETPIDTTIMAKSEVFCVQPEKGGKPARLIV 2580
QY 2581 PPDLGVRVCEKMALYDVVSTLPOAVMGSSYGFQYGFQKQVPEPLVNTWKSXCPMGFSYDT 2640
Db 2581 PPDLGVRVCEKMALYDVVSTLPOAVMGSSYGFQYGFQKQVPEPLVNTWKSXCPMGFSYDT 2640
QY 2641 RCFDSTVTESDIRVEESIYQCCDLAPEARQAIRSLTERLYTGGPLTNSKGQNCGYRRCRA 2700

Db 2641 RCFDSTVTENDIRVEESIYQCCDLAPEARQAIRSLTERLYTGGPLTNSKGQNCGYRRCRA 2700
QY 2701 SGVLTTSCGNLTLCYLKATAACRAAKLQDCTMLVNGDDLVIICBSAGTQEDAAALRAFTE 2760
Db 2701 SGVLTTSCGNLTLCYLKATAACRAAKLQDCTMLVNGDDLVIICBSAGTQEDAAALRAFTE 2760
QY 2761 AMTRYSAAPPDGPQPRYDLLELITSCSSNVSAVHDASGKRVYLLTRDPTTFLARAAWETAR 2820
Db 2761 AMTRYSAAPPDGPQPRYDLLELITSCSSNVSAVHDASGKRVYLLTRDPTTFLARAAWETAR 2820
QY 2821 HTPINSWLGNIMVAPTLWARMILMTHFFPSILLAQOOLEKALDCQIYGACYSIEPDLDPQ 2880
Db 2821 HTPINSWLGNIMVAPTLWARMILMTHFFPSILLAQOOLEKALDCQIYGACYSIEPDLDPQ 2880
QY 2881 IIERLHGLSAFTLHSHYSPGBINRVASCLRKLGVPPLRTWHRARSVRAKLLSOGGRAATC 2940
Db 2881 IIERLHGLSAFTLHSHYSPGBINRVASCLRKLGVPPLRTWHRARSVRAKLLSOGGRAATC 2940
QY 2941 GRYLFWAVRTKLLKLTPIPAASOLDJSGWPFVAGYSGGDIYHSLSRARPRMFPPLCLLLSV 3000
Db 2941 GRYLFWAVRTKLLKLTPIPAASOLDJSGWPFVAGYSGGDIYHSLSRARPRMFPPLCLLLSV 3000
QY 3001 GVGIIYLLPNR 3010
Db 3001 GVGIIYLLPNR 3010

Search completed: April 8, 2004, 10:46:32
Job time : 104 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 14:40:01 ; Search time 35304 Seconds
(without alignments)
11779.867 Million cell updates/sec

Title: US-09-662-454-4
Perfect score: 9595
Sequence: 1 gccagcccccctgatgggggc.....ggcctctctgcagatcatgt 9595

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Result No.	Score	Query Match %	Length	DB	ID	Description
1	9595	100.0	9595	6	AR119832	Sequence
2	9595	100.0	9595	14	AF054247	AR119832 Sequence
3	9518.2	99.2	9595	14	AF054248	AF054247 Hepatitis
4	9473.6	98.7	9596	14	AF054249	AF054248 Hepatitis
5	9381.4	97.8	9580	14	AF054250	AF054249 Hepatitis
6	9327.4	97.2	9460	14	HPCJ491	AF054250 Hepatitis
7	9190.4	95.8	9448	14	HPCJ483	D13558 Hepatitis C
8	8780	91.5	8780	14	AF054255	D13558 Hepatitis C
9	8724	90.9	8780	14	AF054257	AF054255 Hepatitis
10	8703.2	90.7	8780	14	AF054252	AF054257 Hepatitis
11	8703.2	90.7	8780	14	AF054252	AF054252 Hepatitis
12	8684.8	90.5	8779	14	AF054251	AF054252 Hepatitis
13	8658.6	90.2	8781	14	AF054254	AF054251 Hepatitis
14	8599.2	89.6	8780	14	AF054259	AF054254 Hepatitis
15	8592.6	89.6	8780	14	AF054253	AF054259 Hepatitis
16	8577	89.4	8781	14	AF054256	AF054253 Hepatitis
17	8502.6	88.6	9379	14	AF207758	AF054256 Hepatitis
18	8391.8	87.5	9435	14	AB049093	AF207758 Hepatitis
19	8387.2	87.4	9616	14	AB049088	AB049093 Hepatitis
20	8380.2	87.3	9587	14	AB080299	AB049088 Hepatitis
21	8357.4	87.1	9377	14	AF207756	AB080299 Hepatitis
22	8353.8	87.1	9547	14	AB049091	AF207756 Hepatitis
23	8348.2	87.0	9587	14	AF333324	AB049091 Hepatitis
24	8343.8	87.0	9573	14	AB049090	AF333324 Hepatitis
25	8342.8	86.9	9379	14	AF165059	AB049090 Hepatitis
26	8342.8	86.9	9379	14	AF165060	AF165059 Hepatitis
27	8339	86.9	9611	14	AF045702	AF165060 Hepatitis
28	8311.8	86.6	9413	6	E09631	AF045702 Hepatitis
29	8311.8	86.6	9413	6	E09631	E09631 Base sequen
30	8311.8	86.6	9413	6	E09631	E09631 HCV genome
31	8302.2	86.5	9413	6	HPCJ491	D90208 Hepatitis C
32	8299.2	86.5	9456	14	HPCRNA	E08461 Smal-Bsr2 r
33	8272.4	86.2	9379	14	AF165051	D10934 Hepatitis C
34	8271.6	86.2	9375	14	AF207754	AF165051 Hepatitis
35	8268.2	86.2	9379	14	AF207767	AF207754 Hepatitis
36	8262.4	86.1	9377	14	AF207763	AF207767 Hepatitis
37	8261	86.1	9374	14	AF207753	AF207763 Hepatitis
38	8250.2	86.0	9605	6	AX739971	AF207753 Hepatitis
39	8250.2	86.0	9605	14	HGJ238799	AX739971 Sequence
40	8242.6	85.9	9379	14	AF165052	AJ238799 Hepatitis
41	8242.6	85.9	9535	14	D85516	AF165052 Hepatitis
42	8230	85.8	9616	14	AF139594	D85516 Hepatitis C
43	8229.8	85.8	9460	14	AB049097	AF139594 Hepatitis
44	8225.4	85.7	9419	14	AB049098	AB049097 Hepatitis
45	8221.2	85.7	9379	14	AF165057	AB049098 Hepatitis

ALIGNMENTS

RESULT 1	AR119832	Sequence 4 from patent US 6153421.	9595 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	AR119832					
DEFINITION	Sequence 4 from patent US 6153421.					
ACCESSION	AR119832					
VERSION	AR119832.1	GI:14102531				
KEYWORDS	.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 9595)					
TITLE	Yanagi,X., Bukh,J., Emerson,S.U. and Purcell,R.H.					
JOURNAL	Cloned genomes of infectious hepatitis C viruses and uses thereof					
FEATURES	Patent: US 6153421-A 4 28-NOV-2000;					
	Location/Qualifiers					

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

[illegible]

DB 2041 CCCCCGTTAACTCGGGGGGTCGTAACGCGACCTTGATCTGCCCCACGGAATGCTCC 2100
QY 2101 GGAAGCACCCCGAGGCTACTTACACAAAATGCGCTCGGGCCCTGGTTGACACCTAGGT 2160
DB 2101 GGAAGCACCCCGAGGCTACTTACACAAAATGCGCTCGGGCCCTGGTTGACACCTAGGT 2160
QY 2161 GCGTACTAGACTACCCATACAGGCTTTGGCACTTACCCCTGCATCTCTCAATTTTCAATCT 2220
DB 2161 GCGTACTAGACTACCCATACAGGCTTTGGCACTTACCCCTGCATCTCTCAATTTTCAATCT 2220
QY 2221 TTAAGGTTAGGATGTAATGTTGGGGGCGTGAGACAGAGCTCAATGCGCATGCAATGGA 2280
DB 2221 TTAAGGTTAGGATGTAATGTTGGGGGCGTGAGACAGAGCTCAATGCGCATGCAATGGA 2280
QY 2281 CTCGAGGAGAGCGCTCTAACTTTGGAGGACAGGATAGGTGAGAACTCAGCCGCTGCTGC 2340
DB 2281 CTCGAGGAGAGCGCTCTAACTTTGGAGGACAGGATAGGTGAGAACTCAGCCGCTGCTGC 2340
QY 2341 TGTCTACACAGAGTGGGAGATACCTCCCTGTGCTTTACACACCTTACCGGCTTTATCCA 2400
DB 2341 TGTCTACACAGAGTGGGAGATACCTCCCTGTGCTTTACACACCTTACCGGCTTTATCCA 2400
QY 2401 CTGCTTTGATCCATCTCCATCAGAACTCGTGGACGTGCAATACCTGTACGGTGTAGGT 2460
DB 2401 CTGCTTTGATCCATCTCCATCAGAACTCGTGGACGTGCAATACCTGTACGGTGTAGGT 2460
QY 2461 CAGCGTTTCTCTCTTGTCAATCAATGGAGTACATCTGTTGCTTTTCTCTCTCTG 2520
DB 2461 CAGCGTTTCTCTCTTGTCAATCAATGGAGTACATCTGTTGCTTTTCTCTCTCTG 2520
QY 2521 CAGAGCGCGCGTGTGCTGCTTGTGATGATGCTGTGATAGCCACCTTACCGGCTTTATCCA 2580
DB 2521 CAGAGCGCGCGTGTGCTGCTTGTGATGATGCTGTGATAGCCACCTTACCGGCTTTATCCA 2580
QY 2581 CTTAGAGAACTTGTGCTGCTCAATGCGCGCTGCTGCTGCGGAGCGCATGCTTCTCT 2640
DB 2581 CTTAGAGAACTTGTGCTGCTCAATGCGCGCTGCTGCTGCGGAGCGCATGCTTCTCT 2640
QY 2641 CTTTCTCTGTTCTCTGCGCGCTGCTGCTCAATTAAGGCGAGGCTGCTCTGCGGGCG 2700
DB 2641 CTTTCTCTGTTCTCTGCGCGCTGCTGCTCAATTAAGGCGAGGCTGCTCTGCGGGCG 2700
QY 2701 CTTATGCTTTTATGCGCTATGCGCGCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 2760
DB 2701 CTTATGCTTTTATGCGCTATGCGCGCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 2760
QY 2761 CTTAGCGCTTGGACCGGAGATGGCTGATCGTGGGGGCTGGGTTCTTGTAGTCTCG 2820
DB 2761 CTTAGCGCTTGGACCGGAGATGGCTGATCGTGGGGGCTGGGTTCTTGTAGTCTCG 2820
QY 2821 TATTTCTGACCTTGTACCATACTACAAAGTGTCTCTCACTAGGCTCATATGTTGTTAC 2880
DB 2821 TATTTCTGACCTTGTACCATACTACAAAGTGTCTCTCACTAGGCTCATATGTTGTTAC 2880
QY 2881 AATATCTTATACAGAGCGGAGCGACATGCAAGTGTGGTTCGCCCTTCAACGTTTC 2940
DB 2881 AATATCTTATACAGAGCGGAGCGACATGCAAGTGTGGTTCGCCCTTCAACGTTTC 2940
QY 2941 GGGGAGGCGCGATGCGATCTATCTCTTCACTGCTGCGGTTTCACTCAGAGTTAATTTTG 3000
DB 2941 GGGGAGGCGCGATGCGATCTATCTCTTCACTGCTGCGGTTTCACTCAGAGTTAATTTTG 3000
QY 3001 ACATACCAAACTCTCTGCTCGCATACTCGGCCCGCTCATGTGCTTCCAGGCTGGCATAA 3060
DB 3001 ACATACCAAACTCTCTGCTCGCATACTCGGCCCGCTCATGTGCTTCCAGGCTGGCATAA 3060
QY 3061 CGAGATGCGGTAATTTGTCGCGCTCAAGGCTCATCTGTCATGATGTTAGTGGCAA 3120
DB 3061 CGAGATGCGGTAATTTGTCGCGCTCAAGGCTCATCTGTCATGATGTTAGTGGCAA 3120
QY 3121 AAGTCGCGGGGTCATATATGTCAAATGGTCTTCAATGAAGCTGGCGCGCTGACAGGTA 3180

DB 3121 AAGTCGCGGGGTCATATATGTCAAATGGTCTTCAATGAAGTCGCGCGCTGACAGGTA 3180
QY 3181 CQTACGTTTATAAACCATCTTACCCACTCGGGGACTGGGCCACCGGGCTTACAGACC 3240
DB 3181 CQTACGTTTATAAACCATCTTACCCACTCGGGGACTGGGCCACCGGGCTTACAGACC 3240
QY 3241 TTGCGGTGCGGTAGAGCCCGCTCTTCTCGCCATGAGACCAAGGTCTATCACTCGG 3300
DB 3241 TTGCGGTGCGGTAGAGCCCGCTCTTCTCGCCATGAGACCAAGGTCTATCACTCGG 3300
QY 3301 GAGCAGACACCGCTCGGTGCGGACATCACTTTGGGTCTACCCCTCTCGGCCCAAGG 3360
DB 3301 GAGCAGACACCGCTCGGTGCGGACATCACTTTGGGTCTACCCCTCTCGGCCCAAGG 3360
QY 3361 GGAAGAGATATTTTGGGACCGGCTGATAGTCTCGAAGGCAAGGGTGGGACTCTTG 3420
DB 3361 GGAAGAGATATTTTGGGACCGGCTGATAGTCTCGAAGGCAAGGGTGGGACTCTTG 3420
QY 3421 CCCCCATCAAGGCTTACTCCCAACAAACCGGGGCTACTTTGGTTGCTATCACTAGCC 3480
DB 3421 CCCCCATCAAGGCTTACTCCCAACAAACCGGGGCTACTTTGGTTGCTATCACTAGCC 3480
QY 3481 TCACAGGCGGGAAGAAGAACAGGTGCAAGGGAGGTTCAAGTGTCTTACCCCAACAC 3540
DB 3481 TCACAGGCGGGAAGAAGAACAGGTGCAAGGGAGGTTCAAGTGTCTTACCCCAACAC 3540
QY 3541 AATCTTTCTGCGGACCTGCACTACAGGCGGTGCTGCTGACTGTCTACCATGCGCTGG 3600
DB 3541 AATCTTTCTGCGGACCTGCACTACAGGCGGTGCTGCTGACTGTCTACCATGCGCTGG 3600
QY 3601 CGAAGACCTTAGCGGCTCCAAAGGTCCAATCACCCAAATGTACACCAATGTAGACCTGG 3660
DB 3601 CGAAGACCTTAGCGGCTCCAAAGGTCCAATCACCCAAATGTAGACCTGG 3660
QY 3661 ACCTGCTGCGGTGCGAGGCGCCCGCGGGCGGCTCCATGACACCATGCGCTGGCA 3720
DB 3661 ACCTGCTGCGGTGCGAGGCGCCCGCGGGCGGCTCCATGACACCATGCGCTGGCA 3720
QY 3721 GCTCGGACCTTACTTGTGTCAGGACATGTGATGCTATTCGGTGGCGCGGGAGCG 3780
DB 3721 GCTCGGACCTTACTTGTGTCAGGACATGTGATGCTATTCGGTGGCGCGGGAGCG 3780
QY 3781 ACAGCAGGGAAGTCTACTCTCCCCAGGCGCGCTCTCTACTCTGAAAGGCTCTCGG 3840
DB 3781 ACAGCAGGGAAGTCTACTCTCCCCAGGCGCGCTCTCTACTCTGAAAGGCTCTCGG 3840
QY 3841 GTCATTTGCTTTGCGCTCGGGGCGAGTGTGCGGGCTCTTCCGGGCTGCTGTGACACC 3900
DB 3841 GTCATTTGCTTTGCGCTTTGCGGGGCACTGCTGCGGGCTCTTCCGGGCTGCTGTGACACC 3900
QY 3901 GGGGGTTCGGAAGCGGCTGCACTTCTACACCGCTTGAAGTCTATGGAATCTACCATGCG 3960
DB 3901 GGGGGTTCGGAAGCGGCTGCACTTCTACACCGCTTGAAGTCTATGGAATCTACCATGCG 3960
QY 3961 CTCGGTCTTACAGACAACTCAACCCCGCGGCTGTACCCGACAGCAATTCAGTGGCAC 4020
DB 3961 CTCGGTCTTACAGACAACTCAACCCCGCGGCTGTACCCGACAGCAATTCAGTGGCAC 4020
QY 4021 ATCTGACGCTCTACTTGGAGCGGCAAGAGACCAAGTGGCGGCTGCTGTATGACGCC 4080
DB 4021 ATCTGACGCTCTACTTGGAGCGGCAAGAGACCAAGTGGCGGCTGCTGTATGACGCC 4080
QY 4081 AAGGTACAAAGTCTCTGCTGCTGAACCGCTGCTGCGGACCTTAGGCTTGGGGGCT 4140
DB 4081 AAGGTACAAAGTCTCTGCTGCTGAACCGCTGCTGCGGACCTTAGGCTTGGGGGCT 4140
QY 4141 ATATGTCCAAAGGCAACGGTATCGACCTTAAATCAGAACTGGGTGAAGCACTTACCA 4200
DB 4141 ATATGTCCAAAGGCAACGGTATCGACCTTAAATCAGAACTGGGTGAAGCACTTACCA 4200
QY 4201 CCGGCGGCTTCCATTAAGTCTTCCACCTATGCAAGTCTTCTGCGGAGCGGTGGTCTG 4260
DB 4201 CCGGCGGCTTCCATTAAGTCTTCCACCTATGCAAGTCTTCTGCGGAGCGGTGGTCTG 4260

QY 4261 GGGGGCCCTATGACATCATAAATATGTGATGAGTGCACCTCAACTGACTCGACTCACTCT 4320
Db 4261 GGGGGCCCTATGACATCATAAATATGTGATGAGTGCACCTCAACTGACTCGACTCACTCT 4320
QY 4321 TGGGATCGGCAAGTCTTGACCAAGCGGAGACGGCTGAGCGGGCTCGTGTGCTCG 4380
Db 4321 TGGGATCGGCAAGTCTTGACCAAGCGGAGACGGCTGAGCGGGCTCGTGTGCTCG 4380
QY 4381 CCACCGCTACACTCCGGGATCGGTACCGTCCGACACCCCAATATCGAGGAATAGGCC 4440
Db 4381 CCACCGCTACACTCCGGGATCGGTACCGTCCGACACCCCAATATCGAGGAATAGGCC 4440
QY 4441 TGTCCAAATGAGAGATCCCTCTTATGCGAAAGCCATCCCCATGAGGCCATCAAGG 4500
Db 4441 TGTCCAAATGAGAGATCCCTCTTATGCGAAAGCCATCCCCATGAGGCCATCAAGG 4500
QY 4501 GGGGAGGAGATCACTTCTGCAATTCAGAGAAATGTGACGAGTCGCGCAAGC 4560
Db 4501 GGGGAGGAGATCACTTCTGCAATTCAGAGAAATGTGACGAGTCGCGCAAGC 4560
QY 4561 TGACAGGCTCGGACTGAACTGTAGCATATTAACGGGGCTTGTATGTCCGTCATAC 4620
Db 4561 TGACAGGCTCGGACTGAACTGTAGCATATTAACGGGGCTTGTATGTCCGTCATAC 4620
QY 4621 GGCCTATCGGAGACCTCGTGTGCGGCAAGAGAGCTCTAATGACGGGTTCACCGGG 4680
Db 4621 GGCCTATCGGAGACCTCGTGTGCGGCAAGAGAGCTCTAATGACGGGTTCACCGGG 4680
QY 4681 ATTCTGACTCAGTGTACGATGCAATCATATGTCACCCAGACGTGACGTTCAGCTGG 4740
Db 4681 ATTCTGACTCAGTGTACGATGCAATCATATGTCACCCAGACGTGACGTTCAGCTGG 4740
QY 4741 ATCCACCTTACCAATGAGAGACGACGCTGCCCAAGACGCGGTGCGCTCGCAAC 4800
Db 4741 ATCCACCTTACCAATGAGAGACGACGCTGCCCAAGACGCGGTGCGCTCGCAAC 4800
QY 4801 GCGAGGTAGAACTGCGAGGGTAGGATGAGCATCTACAGGTTTGTGACTCCAGAGAAC 4860
Db 4801 GCGAGGTAGAACTGCGAGGGTAGGATGAGCATCTACAGGTTTGTGACTCCAGAGAAC 4860
QY 4861 GGCCTCGGGATGTCATCTTCGCTCGCTGCTGAGTGTGATGACCGGGCTGTGCTT 4920
Db 4861 GGCCTCGGGATGTCATCTTCGCTCGCTGCTGAGTGTGATGACCGGGCTGTGCTT 4920
QY 4921 GGTATGACTACGCGGCTGAGACCTCGGTTAGGTTGCGGGCTTACCTAATAACACAG 4980
Db 4921 GGTATGACTACGCGGCTGAGACCTCGGTTAGGTTGCGGGCTTACCTAATAACACAG 4980
QY 4981 GGTTCGCGCTGCGAGACCATCTGAGATCTGAGATCTGAGGAGCGCTTACAGGCGCTCACCC 5040
Db 4981 GGTTCGCGCTGCGAGACCATCTGAGATCTGAGATCTGAGGAGCGCTTACAGGCGCTCACCC 5040
QY 5041 ACATAGATGCCACTTCTGTCGACAGCTTAAACAGGAGAGACACTTCTTACCTACCTGG 5100
Db 5041 ACATAGATGCCACTTCTGTCGACAGCTTAAACAGGAGAGACACTTCTTACCTACCTGG 5100
QY 5101 TGGCATATCAAGTACAGTGTGCGCAGGCTCAAGCTTCACTCCATCGTGGGACCAAA 5160
Db 5101 TGGCATATCAAGTACAGTGTGCGCAGGCTCAAGCTTCACTCCATCGTGGGACCAAA 5160
QY 5161 TGTGGAAGTGTCTATAGGCTGAAACCTACATGCGAGCGGCCAACCCCTGCTGTATA 5220
Db 5161 TGTGGAAGTGTCTATAGGCTGAAACCTACATGCGAGCGGCCAACCCCTGCTGTATA 5220
QY 5221 GGCTAGAGCGCTCAAAATGAGTCACTCCACACAGCGGCCAACCCCTGCTGTATA 5280
Db 5221 GGCTAGAGCGCTCAAAATGAGTCACTCCACACAGCGGCCAACCCCTGCTGTATA 5280
QY 5281 CATGCATGTGGCTGACCTGGAGTGTCTCATAGCAGCTGGGTCTGCTAGGCGGAGTCC 5340
Db 5281 CATGCATGTGGCTGACCTGGAGTGTCTCATAGCAGCTGGGTCTGCTAGGCGGAGTCC 5340

QY 5341 TTGCAGCTTTGGCCGCACTACTGCTGACGACAGGACAGTGTGCTCATTTGTGGCAGGATCA 5400
Db 5341 TTGCAGCTTTGGCCGCACTACTGCTGACGACAGGACAGTGTGCTCATTTGTGGCAGGATCA 5400
QY 5401 TCTTGTCCGGGAAGCCAGTCTGCTTCCGACAGGGAAGTCTCTTACAGAGTTCGATG 5460
Db 5401 TCTTGTCCGGGAAGCCAGTCTGCTTCCGACAGGGAAGTCTCTTACAGAGTTCGATG 5460
QY 5461 AGATGGAAGTGTGCTCTCAACATTTCTTATCATGAGCAGGGAATGAGTCCCGGAGC 5520
Db 5461 AGATGGAAGTGTGCTCTCAACATTTCTTATCATGAGCAGGGAATGAGTCCCGGAGC 5520
QY 5521 AATTCAAGCAAAAGCGCTCGGTGTGCAAAAGCCACCAAGCAAGAGCGAGCTGCTG 5580
Db 5521 AATTCAAGCAAAAGCGCTCGGTGTGCAAAAGCCACCAAGCAAGAGCGAGCTGCTG 5580
QY 5581 CTCCGCTGTGAGTCCAAAGTGGCGAGCCCTTGAGACTTCTGGCGGAAGCACAATGGA 5640
Db 5581 CTCCGCTGTGAGTCCAAAGTGGCGAGCCCTTGAGACTTCTGGCGGAAGCACAATGGA 5640
QY 5641 ATTTTCATCAGCGGAATACAGTACCTAGAGGCTTATCCACTCTGCTCGGAAACCCCGGA 5700
Db 5641 ATTTTCATCAGCGGAATACAGTACCTAGAGGCTTATCCACTCTGCTCGGAAACCCCGGA 5700
QY 5701 TAGCATCAATTGATGGCAATTTACAGCTTCTATCAGTCCGCTCAGCCCAACCAACCCC 5760
Db 5701 TAGCATCAATTGATGGCAATTTACAGCTTCTATCAGTCCGCTCAGCCCAACCAACCCC 5760
QY 5761 TCTCTTTAAACATCTTGGGGGATGGGTGCTGCCCAACTCGCTTCCCGAGCGCTCGT 5820
Db 5761 TCTCTTTAAACATCTTGGGGGATGGGTGCTGCCCAACTCGCTTCCCGAGCGCTCGT 5820
QY 5821 CAGCTTTCGTGGGCGCGGATCGCCGAGCGGCTGTGGAGCATAGGCTTGGGAAG 5880
Db 5821 CAGCTTTCGTGGGCGCGGATCGCCGAGCGGCTGTGGAGCATAGGCTTGGGAAG 5880
QY 5881 TGCTCTGTGACATCTGCGCGGCTATGGGCGAGGGTAGCGGCGCACTCGTGGCCTTTA 5940
Db 5881 TGCTCTGTGACATCTGCGCGGCTATGGGCGAGGGTAGCGGCGCACTCGTGGCCTTTA 5940
QY 5941 AGGTCTAGAGCGGCGAGTCCCTTCCACCGAGGACCTTGTCAACTTACTCCCTGCCATCC 6000
Db 5941 AGGTCTAGAGCGGCGAGTCCCTTCCACCGAGGACCTTGTCAACTTACTCCCTGCCATCC 6000
QY 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060
Db 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060
QY 6061 GCGCGGAGAGGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6120
Db 6061 GCGCGGAGAGGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6120
QY 6121 ACCAGCTCTCCCTACGACTATGTGCTGAGAGCGAGCTGCGAGCACTGCTGCTGCTGCTGCT 6180
Db 6121 ACCAGCTCTCCCTACGACTATGTGCTGAGAGCGAGCTGCGAGCACTGCTGCTGCTGCTGCT 6180
QY 6181 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6240
Db 6181 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6240
QY 6241 ACTGCTCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6300
Db 6241 ACTGCTCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6300
QY 6301 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6360
Db 6301 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6360
QY 6361 CTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6420
Db 6361 CTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6420
QY 6421 CCACCTGCCATGCGGAGGACAGATCGCGGACATGTCAAAACCGCTTCCATGAGGATCG 6480

QY	8641	CCCCCCCGGGATCGCCCCCAACCGAATACGACCTGGAGCTGATAACAATCATGTTCTT	8700
Db	8641	CCCCCCCGGGATCGCCCCCAACCGAATACGACCTGGAGCTGATAACAATCATGTTCTT	8700
QY	8701	CGAATGTCAGTCGGCGCAGATGATCTGGGAAAGGTTACTACCTCACCCGTCACC	8760
Db	8701	CGAATGTCAGTCGGCGCAGATGATCTGGGAAAGGTTACTACCTCACCCGTCACC	8760
QY	8761	CCACCACCCCTTGCACGGCTGCGTGGGAGACAGTACACACCTCCCAATCACTCTT	8820
Db	8761	CCACCACCCCTTGCACGGCTGCGTGGGAGACAGTACACACCTCCCAATCACTCTT	8820
QY	8821	GGTAGGCAATATCATATGATGCGCCACCTATGGCAAGGATGATTCGTGATGACTC	8880
Db	8821	GGTAGGCAATATCATATGATGCGCCACCTATGGCAAGGATGATTCGTGATGACTC	8880
QY	8881	ACTTTTCTCCACCTCTTAGCTCAGAGCACTTGAAGCCCTGGATTTGTGATCT	8940
Db	8881	ACTTTTCTCCACCTCTTAGCTCAGAGCACTTGAAGCCCTGGATTTGTGATCT	8940
QY	8941	ACGGGCTTGTACTCCATTTAGCCACTTGACCTACCTCAGATCATTAACGACTCCATG	9000
Db	8941	ACGGGCTTGTACTCCATTTAGCCACTTGACCTACCTCAGATCATTAACGACTCCATG	9000
QY	9001	GTCTTAGGCAATTTACACTCCAGTACTCTCCAGGTGAGATCAATAGGTTGGCTTCA	9060
Db	9001	GTCTTAGGCAATTTACACTCCAGTACTCTCCAGGTGAGATCAATAGGTTGGCTTCA	9060
QY	9061	GCTCAGGAACTTGGGTTACCACTTGGGTAACCTTGGGTAACCTTGGGTAACCTTGGG	9120
Db	9061	GCTCAGGAACTTGGGTTACCACTTGGGTAACCTTGGGTAACCTTGGGTAACCTTGGG	9120
QY	9121	CGCTAAGCTACTGTCACAGGGGGAGGGCCGACCTTGTGGCAGATACCTTCTTAAC	9180
Db	9121	CGCTAAGCTACTGTCACAGGGGGAGGGCCGACCTTGTGGCAGATACCTTCTTAAC	9180
QY	9181	GGGCAATAGGACCAAGCTTAACTCCTCCTAACTCCCGGCGGCTGCTGCTGCTGCT	9240
Db	9181	GGGCAATAGGACCAAGCTTAACTCCTCCTAACTCCCGGCGGCTGCTGCTGCTGCT	9240
QY	9241	CTGGCTGGTTCTGCTGCTGCTTACAGGGGGAGGGCCGACCTTGTGGCAGATACCT	9300
Db	9241	CTGGCTGGTTCTGCTGCTGCTTACAGGGGGAGGGCCGACCTTGTGGCAGATACCT	9300
QY	9301	GACCCGCTGCTGCTGCTGCTTACAGGGGGAGGGCCGACCTTGTGGCAGATACCT	9360
Db	9301	GACCCGCTGCTGCTGCTGCTTACAGGGGGAGGGCCGACCTTGTGGCAGATACCT	9360
QY	9361	TCGCCAAGGATGAGGGGAGCTAACCTACCTGCTGCTGCTGCTGCTGCTGCTGCT	9420
Db	9361	TCGCCAAGGATGAGGGGAGCTAACCTACCTGCTGCTGCTGCTGCTGCTGCTGCT	9420
QY	9421	TT	9480
Db	9421	TT	9480
QY	9481	TTTTTCCCTTCTTAATGCTGCTCACTTACCTGCTGCTGCTGCTGCTGCTGCTGCT	9540
Db	9481	TTTTTCCCTTCTTAATGCTGCTCACTTACCTGCTGCTGCTGCTGCTGCTGCTGCT	9540
QY	9541	CGGTGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	9595
Db	9541	CGGTGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	9595

RESULT 2
 LOCUS AF054247
 DEFINITION Hepatitis C virus HC-J4, pCV-J4L6s, complete genome.
 ACCESSION AF054247
 VERSION AF054247.1 GI:3098632
 KEYWORDS

SOURCE ORGANISM
 Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
 REFERENCE 1 (bases 1 to 9595)
 AUTHORS Yanagi.M., St Claire.M., Shapiro.M., Emerson.S.U., Purcell.R.H. and Bukh.J.
 TITLE Transcripts of a chimeric cDNA clone of hepatitis C virus genotype 1b are infectious in vivo
 JOURNAL Virology 244 (1), 161-172 (1998)
 MEDLINE 95240944
 PUBMED 9581788
 REFERENCE 2 (bases 1 to 9595)
 AUTHORS Yanagi.M. and Bukh.J.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAR-1998) Hepatitis Viruses Section, LID, NIAID, NIH, 7 Center Dr., Building 7, Room 201, Bethesda, MD 20892, USA

CDS

1b are infectious in vivo

1. .9595

/organism="Hepatitis C virus"

/mol_type="genomic RNA"

/strain="HC-J4"

/isolate="pCV-J4L6s"

/db_xref="taxon:11103"

/note="infectious full-length cDNA"

342..9374

/codon_start=1

/product="polyprotein"

/protein_id="AAC15722.1"

/db_xref="GI:3098633"

/translation="MSTNRPQRTKENTNRNPDQVFPGGQYGVGVYLLPRGPRLL

GVRATKASRSQPRRQPTKARRPEGRGAWOPGYPWLPYNEFGLGWAGLLSBRG

SRPWSGNTDPRRRSNGKVIDTLTCGADLMGYIPLVAGPLCGAARALAHGVLE

GVNATGNTLPCGSPSIFILLALSLTTFASAYEVNRVSGIYHVINDSNSISVFAAD

VIMTPGCVPCQSGNSRWALPTLAARNSVPTTIRRHVDLILVALLPAGVDGTHTRVA

SQLRIIPQAVVSVAGAWGLAYYSVMVKNWLVILVALLPAGVDGTHTRVA

GHTSGTSLFSSGASQKIQLVNTGSHNINRTALNCNLSLQCTFPFALFAHFNFS

CGPRMASCRIPDFAOGWGPITVTKPNSDQRPVCHYAPRGVGPVPSQVCGPVVC

TPSPVVGTVDRSGVPTYSWGENETVMLNTRPFGWPGTWNSTGPTCCG

PPNIGVGNRLLICPTDCPRKPEANYTKCGSPWLTPLCLVDYVRLWHYPTCLNP

SIFKRMVGVGVEHRLNAACNWTGRACNLEDRDRSELSPULLSTWQLIPCAFTL

PAUSTGLTHQNTVDVYLYGSAFVSFAIKWEYILLFLLADARVACLMWMLL

IAEAALNLNLVNLAAVAGAHGILSFVFAACAWYIKRGLAPGAAYVAFVGMWLL

LLALPPRALVDEMAACGGAVLGLVFLITSPYKVFPLIWMLOVFTIETAEAH

MOVVPLNVGGRDAIILLTCAVHPELIDITKLALLGLPLMLVLAQITRVPVVR

AOGLIRACMLVRKAGSHYQVMPKLGALITGYVYNHLEFLDRWAHAGLRDLAVVE

PVPSMETKITWGAADTAACGDIILGLPVSARKKSEIFLGPADSLGGQRLLIAPIT

AYSQTRGVLCIITSUTGRDQVEGEVQVSTATQSLFATCINGCYWTYHGAGSK

TLAGKGFITQMYTNVDLDLWQAPPGARSMTPCSGSSDLVLTVRHADVIVRRRG

DSRGLSPRPVSYLKGSGPLPCSHVGVFRAAVCTRGVAKAVDFIVPESMETT

MRSFVTDNSTPAPVOTFOVHLHAPTGSKSTKVPAAVAAQYKVLINPVAATL

GEGAYMSKHGIDNIRGTITTGSIYSYTKGLADGGSGGAYDILIIIDECBS

TDSTIIIGITVLQDAETAGRLVVLATATPGSVTVPHNIEIIGUNNGEIPFYK

AIPIEALKGRHLIFHSKCKDBLAALITGLINAVAYYGLDVSVIPPIGIDVVVA

TDALMTGDFDVIDCNCVTQTVDFSLDPTFTIETTVQDADVSRRGTGSG

RSGLIRFVTPGERSGFMDSVLCYDAGCAWYELTPARTVRLRAQAPPPSWQVQC

DHLEFESVFTGLTHIDHRLSOTKQAGDNFVILVQATVCAQAQAPPSWQVQC

LIRLKP2TLHSEPTLLYRLGAVQNEVILITIKIMACMSADLEVTSTWLVGVULA

ALAXCLITSGVIVGRIILSGPAPVDPREVLYQEPDEECASQYLFIEQGLQALAE

QKQALGLQATKQAAAPVPSKRALETFWAKMWNFISSQIYLGALSTLPON

PAIASMAFTASITSLTNTLNLGVAQAAPPSSAFAFVAGLAGAVGSI

GLGSLVDILAGAGVAGALVAFKVMGSEVSTEDLNLPAISLPAISLPAISLPAIS

LIRHVGPGVAGVQMMRLIAFASRGNHVSPTHYVPESDAARVTOISLITLQALK

RLHOWINEDCSTPCSGSLRDVMDWICTVLTDFKWLQSKLLPRLPGVPLSCQGVK

GVNRGDGIMQITCEPCQAIGHVGNMRIVGPTCNCTWEGFTPINAYITGPTPSP

APNYSRALVRADAEYVEVTRVGFHYVTGTTDNDKPCQCPQFAPEFTFVRLHR

YAPACKLLEEDVTFQVNLQYLVGSQLPCEPEPDVTLTSLMTDPGSHIYAEAKRL

ARGSPSIIAGSSASQSLKATFTIIRKSRPSALPIWARPDYNPILLESWKDPDY

KVILDSFEPHLAGEDRRSTVAATFTIIRKSRPSALPIWARPDYNPILLESWKDPDY

VPPVHGCPLPPTKAPPIPPPRKRTVLTSTVSSALAEATKTPCSSSSGSAVDSGT

ATALPDUAASDGGKSDVESISGMPPEGPDPDLSDGSWSTVSEASVJVCSSMS

YTWGALITPCAAEESKLPINPLSNLSLRHNNVYATTSRSLRQKVVTFDRLQVLD

1921 A TAGTGGGGGAGATGAGACAGACAGTGTATGCTCTCTCAACACACGCGTCCGCCACAAG 1980
1981 GCAACTGGTTCCGCTGTATCATGATGAATAGTACTGGTTCTACTAAGACGTGCGAGGTC 2040
1981 GCACTGGTTCCGCTGTATCATGATGAATAGTACTGGTTCTACTAAGACGTGCGAGGTC 2040
2041 CCCCGTGAACATTCGGGGGGTCCGTPAACCGCACCTTGATCTGCCCCACGACTGCTCC 2100
2041 CCCCGTGAACATTCGGGGGGTCCGTPAACCGCACCTTGATCTGCCCCACGACTGCTCC 2100
2101 GGAAGCACCCCGAGGCTACTTACAAAATGTGGCTCGGGGCGCTGGTGAACCTAGGT 2160
2101 GGAAGCACCCCGAGGCTACTTACAAAATGTGGCTCGGGGCGCTGGTGAACCTAGGT 2160
2161 GCTAGTAGACTACCCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTCCATCT 2220
2161 GCTAGTAGACTACCCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTCCATCT 2220
2221 TTAAAGTTAGGATGTATGTGGGGGGTGGAGACAGGCTCAAGTCCGATGCAATTTGA 2280
2221 TTAAAGTTAGGATGTATGTGGGGGGTGGAGACAGGCTCAAGTCCGATGCAATTTGA 2280
2281 CTCGAGGAGCGCTGTAACTTTGGAGGACAGGCTAGGTCAGAACTCAGCCCGCTGCTGC 2340
2281 CTCGAGGAGCGCTGTAACTTTGGAGGACAGGCTAGGTCAGAACTCAGCCCGCTGCTGC 2340
2341 TGTCTACAAAGAGTGGCAGACTACTGCCCTGTGCTTTTACACCCCTACCGGCTTTATCCA 2400
2341 TGTCTACAAAGAGTGGCAGACTACTGCCCTGTGCTTTTACACCCCTACCGGCTTTATCCA 2400
2401 CTGGTTTGAATCATCTCATAGACATCTGTGGACGTGGCAATPACCTGTACGGTGTAGGCT 2460
2401 CTGGTTTGAATCATCTCATAGACATCTGTGGACGTGGCAATPACCTGTGTACGGTGTAGGCT 2460
2461 CAGCCCTTTGTCTCCCTTTGCAATCAAAATGGAGTACATCTCTGTGTTTCTCTCTCTCTGG 2520
2461 CAGCCCTTTGTCTCCCTTTGCAATCAAAATGGAGTACATCTCTGTGTTTCTCTCTCTCTGG 2520
2521 CAGACCGCGCTGTGTGCTGCTTTGTGATGATGCTGTGATAGCCAGGCTGAGGCGG 2580
2521 CAGACCGCGCTGTGTGCTGCTTTGTGATGATGCTGTGATAGCCAGGCTGAGGCGG 2580
2581 CATTAGAGAACTTGTGTGCTCAATGGCGCTCGTGGCCGCGAGCGATGTATCTCT 2640
2581 CATTAGAGAACTTGTGTGCTCAATGGCGCTCGTGGCCGCGAGCGATGTATCTCT 2640
2641 CTTTCTTGTGCTCTCTCGCGCTGCTGATCAATTAAGGCGAGGCTGCTCTCTGGGCGG 2700
2641 CTTTCTTGTGCTCTCTCGCGCTGCTGATCAATTAAGGCGAGGCTGCTCTCTGGGCGG 2700
2701 CGTATGCTTTTATGGCTATGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
2701 CGTATGCTTTTATGGCTATGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
2761 CTTAGCCCTTGGACCGGAGATGGTGCATCTGCGGGGGTGGCTTTCTGTAGTCTGG 2820
2761 CTTAGCCCTTGGACCGGAGATGGTGCATCTGCGGGGGTGGCTTTCTGTAGTCTGG 2820
2821 TATTCTTACCTTGTACCACTACTACAAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
2821 TATTCTTACCTTGTACCACTACTACAAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
2881 AATACCTTTATCAGACAGCGAGGCGCATGCAAGTGTGGTTCGCCCTCTCAAGCTTC 2940
2881 AATACCTTTATCAGACAGCGAGGCGCATGCAAGTGTGGTTCGCCCTCTCAAGCTTC 2940
2941 GGGAGGCGCGATGCCATCATCT 3000
2941 GGGAGGCGCGATGCCATCATCT 3000
3001 ACATCAAACTCTGTGTGCCATPACTCGGCCCGCTCATGGTGTCTCAGGCTGGCATAA 3060
3001 ACATCAAACTCTGTGTGCCATPACTCGGCCCGCTCATGGTGTCTCAGGCTGGCATAA 3060

3061 CGAGATGCCCTACTTCTGTCGCGCTCAAGGCTCATTCGTGCATGATCTTAGTGCAG 3120
3061 CGAGATGCCCTACTTCTGTCGCGCTCAAGGCTCATTCGTGCATGATCTTAGTGCAG 3120
3121 AAGTCCGCGGGGCTATTATGTCCAAATGTCTTTCATGAAGCTGGCGCTGACAGGTA 3180
3121 AAGTCCGCGGGGCTATTATGTCCAAATGTCTTTCATGAAGCTGGCGCTGACAGGTA 3180
3181 GGTAGCTTTATAACCACTTTACCCCACTGCGGGACTGGGCCACGCGGGCTTACGAGAC 3240
3181 GGTAGCTTTATAACCACTTTACCCCACTGCGGGACTGGGCCACGCGGGCTTACGAGAC 3240
3241 TTGCGTGGCGGTAGAGCCGCTGCTTCTCCGCCATGGAGACCAAGTCACTACCTGGG 3300
3241 TTGCGTGGCGGTAGAGCCGCTGCTTCTCCGCCATGGAGACCAAGTCACTACCTGGG 3300
3301 GAGCAGACACCGCTCGCTGTGGGACATCATCTTTGGGTCTACCCGTCTCCGCCCGAAGG 3360
3301 GAGCAGACACCGCTCGCTGTGGGACATCATCTTTGGGTCTACCCGTCTCCGCCCGAAGG 3360
3361 GGAAGGAGATATTTTGGGACCGCTGATAGTCTCGAAGGCAAGGTTGGGACTCTCTTG 3420
3361 GGAAGGAGATATTTTGGGACCGCTGATAGTCTCGAAGGCAAGGTTGGGACTCTCTTG 3420
3421 CGCCCATCAGCGCTACTCTCCCAACAGCGGGGCTACTTTGGTTGCATCATCTACCTAGC 3480
3421 CGCCCATCAGCGCTACTCTCCCAACAGCGGGGCTACTTTGGTTGCATCATCTACCTAGC 3480
3481 TCACAGGCGGCAACAAGAACAGGTCCAAATCACCCAAATGTACCAATGTACACCTGG 3540
3481 TCACAGGCGGCAACAAGAACAGGTCCAAATCACCCAAATGTACCAATGTACACCTGG 3540
3541 AATCTTCTTGGCGACCTGATCAACCGCGTGTCTGCTGCTACTACATGGCGCTGGCT 3600
3541 AATCTTCTTGGCGACCTGATCAACCGCGTGTCTGCTGCTACTACATGGCGCTGGCT 3600
3601 CGAAGACCTTAGCCGCTCCAAATAGGTCCAAATCACCCAAATGTACCAATGTACACCTGG 3660
3601 CGAAGACCTTAGCCGCTCCAAATAGGTCCAAATCACCCAAATGTACCAATGTACACCTGG 3660
3661 ACCTCGTGGTGGAGCGGCGGCGGCGGCGGCGCTCCATGACACCATGAGCTGTGGCA 3720
3661 ACCTCGTGGTGGAGCGGCGGCGGCGGCGGCGCTCCATGACACCATGAGCTGTGGCA 3720
3721 GCTCGGACCTTACTTGGTCAAGAGACATGCTGATGTCTTTCGCTGCGCGGAGGCG 3780
3721 GCTCGGACCTTACTTGGTCAAGAGACATGCTGATGTCTTTCGCTGCGGAGGCG 3780
3781 ACACGAGGGAAGTCTACTCTCCCGAGGCGCTCTCTTACCTGAAAGGCTCTCTCGGCTG 3840
3781 ACACGAGGGAAGTCTACTCTCCCGAGGCGCTCTCTTACCTGAAAGGCTCTCTCGGCTG 3840
3841 GTCCATTTGCTTTCGCTTCCGGGACGCTGCGGCGCTCTCCGGCTGCTGTGTGACACCC 3900
3841 GTCCATTTGCTTTCGCTTCCGGGACGCTGCGGCGCTCTCCGGCTGCTGTGTGACACCC 3900
3901 GGGGGTTCGCAAGCGGCTGCACTTCAATCCCGTTGAGTCTATGGAAGTACCATGCGGT 3960
3901 GGGGGTTCGCAAGCGGCTGCACTTCAATCCCGTTGAGTCTATGGAAGTACCATGCGGT 3960
3961 CTCGGTCTTACAGACAACTCAACCCCGGCTGTACCGCAGACATTCCTCAAGTGGCAC 4020
3961 CTCGGTCTTACAGACAACTCAACCCCGGCTGTACCGCAGACATTCCTCAAGTGGCAC 4020
4021 ATCTGCAAGCTCTTACTGGCAGGCGAGACACCAAGTGGCGGCTGCTGTATGACGCC 4080
4021 ATCTGCAAGCTCTTACTGGCAGGCGAGACACCAAGTGGCGGCTGCTGTATGACGCC 4080
4081 AAGGTTACAAAGTGTCTCTGCTGAAACCGGCTGCGGCACTTACGAGGTTTGGGCGCT 4140
4081 AAGGTTACAAAGTGTCTCTGCTGAAACCGGCTGCGGCACTTACGAGGTTTGGGCGCT 4140

QY 4141 ATATGTCGAGGACACGCTATCGACCTTAACATCAGAACTGGGGTAAGAACCAATTACCA 4200
 Db 4141 ATATGTCGAGGACACGCTATCGACCTTAACATCAGAACTGGGGTAAGAACCAATTACCA 4200
 QY 4201 CGGGCGGCTCCATTACGTAATCACTTACCTATGGGAAGTTCTTGGCGACGGTGGCTGTCTG 4260
 Db 4201 CGGGCGGCTCCATTACGTAATCACTTACCTATGGGAAGTTCTTGGCGACGGTGGCTGTCTG 4260
 QY 4261 GGGCGGCTTATGACATCATAAATATGATGAGTGGCCACTCAACTGACTCGACTPACCATCT 4320
 Db 4261 GGGCGGCTTATGACATCATAAATATGATGAGTGGCCACTCAACTGACTCGACTPACCATCT 4320
 QY 4321 TGGGCTATCGGACAGTTCCTGGACCAAGCGGAGACGGCTGGAGCGCGCTCGTCTGCTCG 4380
 Db 4321 TGGGCTATCGGACAGTTCCTGGACCAAGCGGAGACGGCTGGAGCGCGCTCGTCTGCTCG 4380
 QY 4381 CCACCGCTACACCTCCGGGATCGGTTACCGTGGCCACACCCCAATATCGAGGAATAAGGCC 4440
 Db 4381 CCACCGCTACACCTCCGGGATCGGTTACCGTGGCCACACCCCAATATCGAGGAATAAGGCC 4440
 QY 4441 TGTCCAAATGAGAGATCCCTTCTATGCGCAAAAGCCATCCCATTTGAGGCCATCAAGG 4500
 Db 4441 TGTCCAAATGAGAGATCCCTTCTATGCGCAAAAGCCATCCCATTTGAGGCCATCAAGG 4500
 QY 4501 GGGGAGGACATCTCATTTCTGCCATTCAGAAAGAAATGTGACGAGCTCGCGGAAAGC 4560
 Db 4501 GGGGAGGACATCTCATTTCTGCCATTCAGAAAGAAATGTGACGAGCTCGCGGAAAGC 4560
 QY 4561 TGACAGGCTCGGACTGAACGCTGTAGATATATTACCGGGCCCTTGATGTGCTGCTATAC 4620
 Db 4561 TGACAGGCTCGGACTGAACGCTGTAGATATATTACCGGGCCCTTGATGTGCTGCTATAC 4620
 QY 4621 CGCCTATCGGAGACGCTGTGTGCTGGCAACAGACGCTCTAATGACGGGTTTCAACGGGG 4680
 Db 4621 CGCCTATCGGAGACGCTGTGTGCTGGCAACAGACGCTCTAATGACGGGTTTCAACGGGG 4680
 QY 4681 ATTTGACTGATGATCGATCGATCAATATGATGTGTCACCCAGACGTCGACTTCAGCTTGG 4740
 Db 4681 ATTTTGAATGATGATCGATCGATCAATATGATGTGTCACCCAGACGTCGACTTCAGCTTGG 4740
 QY 4741 ATCCCACTTACCAATGAGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4800
 Db 4741 ATCCCACTTACCAATGAGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4800
 QY 4801 GCGAGGTAGAACTGCGAGGGTAGGAGTGGCATCTACAGGTTTGTGACTCCAGGAGAAC 4860
 Db 4801 GCGAGGTAGAACTGCGAGGGTAGGAGTGGCATCTACAGGTTTGTGACTCCAGGAGAAC 4860
 QY 4861 GGGCTCGGGGATGTCGATTCCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
 Db 4861 GGGCTCGGGGATGTCGATTCCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
 QY 4921 GGTATGAGCTACGCGCTGAGACCTCGGTTAGGTTGGGGCTTACCTAATATACCCAG 4980
 Db 4921 GGTATGAGCTACGCGCTGAGACCTCGGTTAGGTTGGGGCTTACCTAATATACCCAG 4980
 QY 4981 GGTTCGCTGTCGAGGACCACTTGGAGTTCTGGAGAGCGCTTTCACAGGCTTCAACC 5040
 Db 4981 GGTTCGCTGTCGAGGACCACTTGGAGTTCTGGAGAGCGCTTTCACAGGCTTCAACC 5040
 QY 5041 ACATAGATCCGACCTTCTGTCGAGCTTACCTTACCTGACGCGGCAACACCCCTGCTGTATA 5100
 Db 5041 ACATAGATCCGACCTTCTGTCGAGCTTACCTTACCTGACGCGGCAACACCCCTGCTGTATA 5100
 QY 5101 TGGCATATCAAGTACAGTGTGCGGACGAGGCTCAAGCTCCATCCATCGTGGGCCAAA 5160
 Db 5101 TGGCATATCAAGTACAGTGTGCGGACGAGGCTCAAGCTCCATCCATCGTGGGCCAAA 5160
 QY 5161 TGTGAAGTGTCTCATACGGCTGAACCTTACCTGACGCGGCAACACCCCTGCTGTATA 5220
 Db 5161 TGTGAAGTGTCTCATACGGCTGAACCTTACCTGACGCGGCAACACCCCTGCTGTATA 5220
 QY 5221 GGCTAGGACCGCTCCAAATGAGGTGATCTCCACACACCCCAATAAATAATCATATGG 5280

Db 5221 GGCTAGGACCGCTCCAAATGAGGTGATCTCCACACACCCCAATAAATAATCATATGG 5280
 QY 5281 CATGATGCTGCGCTGACCTGGAGGTGCTCATAGACACCTGGGTGCTGTAGCGGAGTCC 5340
 Db 5281 CATGATGCTGCGCTGACCTGGAGGTGCTCATAGACACCTGGGTGCTGTAGCGGAGTCC 5340
 QY 5341 TTGCAAGCTTTGGCCGCTACTGCTTCAGACAGGCAAGTGTGGTCAATGTTGGGCAAGATCA 5400
 Db 5341 TTGCAAGCTTTGGCCGCTACTGCTTCAGACAGGCAAGTGTGGTCAATGTTGGGCAAGATCA 5400
 QY 5401 TCTTGTCCGGGAAGCCAGCTGTCTTCCCGACAGGGAAGTCTCTACAGGAGTTTCGATG 5460
 Db 5401 TCTTGTCCGGGAAGCCAGCTGTCTTCCCGACAGGGAAGTCTCTACAGGAGTTTCGATG 5460
 QY 5461 AGATGGAAGTGTGCTCAACACTTCTTATCATGAGCAGGGAATGACGCTTCGCGGAGC 5520
 Db 5461 AGATGGAAGTGTGCTCAACACTTCTTATCATGAGCAGGGAATGACGCTTCGCGGAGC 5520
 QY 5521 RAATTCAGCAAAAGCGCTCGGTTTGTGCAAAAGCGCCACCAAGCAAGCGAGGCTGCTG 5580
 Db 5521 RAATTCAGCAAAAGCGCTCGGTTTGTGCAAAAGCGCCACCAAGCAAGCGAGGCTGCTG 5580
 QY 5581 CTCCGCTGTGAGTCCAAAGTGGCGAGCCCTTGAGACCTTCTGGCGGAAGCAATGTGGA 5640
 Db 5581 CTCCGCTGTGAGTCCAAAGTGGCGAGCCCTTGAGACCTTCTGGCGGAAGCAATGTGGA 5640
 QY 5641 ATTTTCATCAGCGGAATACAGTACCTTAGCAGGCTTATCCACTCTGCTGGAACCCCGGGA 5700
 Db 5641 ATTTTCATCAGCGGAATACAGTACCTTAGCAGGCTTATCCACTCTGCTGGAACCCCGGGA 5700
 QY 5701 TAGCATCATTTGATGCACTTTACAGCTTCTATCACTAGCCGCTCACCAACCCCAACACCC 5760
 Db 5701 TAGCATCATTTGATGCACTTTACAGCTTCTATCACTAGCCGCTCACCAACCCCAACACCC 5760
 QY 5761 TCTCTTTTAAACATCTTGGGGGATGGGTGGCTGCGCCAACTCGCTCTCCACGCTGCTG 5820
 Db 5761 TCTCTTTTAAACATCTTGGGGGATGGGTGGCTGCGCCAACTCGCTCTCCACGCTGCTG 5820
 QY 5821 CAGCTTTCTGGGCGCCGCGCATTCGCGGAGCGGCTGTTGGCAGCATAGGCTTGGGAAG 5880
 Db 5821 CAGCTTTCTGGGCGCCGCGCATTCGCGGAGCGGCTGTTGGCAGCATAGGCTTGGGAAG 5880
 QY 5881 TGCTCGTGAACATCTTGGCGGCTATGGGCGAGGGGTAGCGCGGCACTCGTGGGCTTTA 5940
 Db 5881 TGCTCGTGAACATCTTGGCGGCTATGGGCGAGGGGTAGCGCGGCACTCGTGGGCTTTA 5940
 QY 5941 AGGTCAATGAGCGCGGAGTGGCTTCCACCGAGGACCTTGGTCAACTTATCTGCTGCACTCC 6000
 Db 5941 AGGTCAATGAGCGCGGAGTGGCTTCCACCGAGGACCTTGGTCAACTTATCTGCTGCACTCC 6000
 QY 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060
 Db 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060
 QY 6061 GCGCGGAGAGGGGGCTGTCAGTGAATGAACCGGCTCATAGCGTTCTGCTTCCGCGGCTA 6120
 Db 6061 GCGCGGAGAGGGGGCTGTCAGTGAATGAACCGGCTCATAGCGTTCTGCTTCCGCGGCTA 6120
 QY 6121 ACCAGCTCTCCCTTACGCACTATGTCCTGAGGACGCTGACAGCTGCTGCTGCTGCTGCTGCT 6180
 Db 6121 ACCAGCTCTCCCTTACGCACTATGTCCTGAGGACGCTGACAGCTGCTGCTGCTGCTGCTGCT 6180
 QY 6181 TCTCTCTGAGCTTACCATCTCACTCACTGCTGAGCGGCTCCACAGTGGATTAATGAGG 6240
 Db 6181 TCTCTCTGAGCTTACCATCTCACTCACTGCTGAGCGGCTCCACAGTGGATTAATGAGG 6240
 QY 6241 ACTGCTTACGCCATGCTTCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6300
 Db 6241 ACTGCTTACGCCATGCTTCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6300
 QY 6301 TGTGACTGATTTCAAGACCTTGGCTCCAGTCCAACTCTCTCCGCGGTTACCGGAGTCC 6360

Db 6301 TCTTGACTGACTTCAAGACCTGECTCCAGTCCAAAACTCTCTGCGCGGTCTAACCGGAGTCC 6360
Qy 6361 CTTTCTCTGTCATGCGAAACGCGGGTCAAGGGAGTCTGGCGGGGAGCGGCATCATGCGAAA 6420
Db 6361 CTTTCTCTGTCATGCGAAACGCGGGTCAAGGGAGTCTGGCGGGGAGCGGCATCATGCGAAA 6420
Qy 6421 CCACCTTGCCCATGCGGAGCACAGATGCGCGGACATGTCAAAAAGGTTCCCATGAGATCG 6480
Db 6421 CCACCTTGCCCATGCGGAGCACAGATGCGCGGACATGTCAAAAAGGTTCCCATGAGATCG 6480
Qy 6481 TAGGGCCTAGAACCTTCGACGACACAGTGTGSCACGGAAAGTTCGCCATCAACGCAATACACA 6540
Db 6481 TAGGGCCTAGAACCTTCGACGACACAGTGTGSCACGGAAAGTTCGCCATCAACGCAATACACA 6540
Qy 6541 CCGGACCTTGTGCACACCTCCCGCGGCGCAACTATTCCAGGGCGCTATGCGCGGTGGGTG 6600
Db 6541 CCGGACCTTGTGCACACCTCCCGCGGCGCAACTATTCCAGGGCGCTATGCGCGGTGGGTG 6600
Qy 6601 CTGAGGAGTACGTGAGAGTTACGCGTGTGGGGGAATTCACCTACGCTGACGGGCATGACCA 6660
Db 6601 CTGAGGAGTACGTGAGAGTTACGCGTGTGGGGGAATTCACCTACGCTGACGGGCATGACCA 6660
Qy 6661 CTGACAAAGTAAAGTGCATGCCAGGTTCCGGGCGGTGCAAACTCTTCTACGGGAGGACGTCA 6720
Db 6661 CTGACAAAGTAAAGTGCATGCCAGGTTCCGGGCGGTGCAAACTCTTCTACGGGAGGACGTCA 6720
Qy 6721 GAGTCCGGTTGCACAGGTACGCTCCGGGCGTCAAACTCTTCTACGGGAGGACGTCA 6780
Db 6721 GAGTCCGGTTGCACAGGTACGCTCCGGGCGTCAAACTCTTCTACGGGAGGACGTCA 6780
Qy 6781 TCCAGGTGGGTCTCAACCAATCTTGTTGGTGGGTGGAGTCCCATGCGAGGCCGAACCGG 6840
Db 6781 TCCAGGTGGGTCTCAACCAATCTTGTTGGTGGGTGGAGTCCCATGCGAGGCCGAACCGG 6840
Qy 6841 AGCTAACAGTCTTACTTCCATGCTCACGATCCGATCCCACTTACACAGAGCGGCTA 6900
Db 6841 AGCTAACAGTCTTACTTCCATGCTCACGATCCGATCCCACTTACACAGAGCGGCTA 6900
Qy 6901 AGCTAGGCTGCTAGAGGTTCTCCCGCTCTTTAGCCAGCTCATCAGTACGCCAGTTGT 6960
Db 6901 AGCTAGGCTGCTAGAGGTTCTCCCGCTCTTTAGCCAGCTCATCAGTACGCCAGTTGT 6960
Qy 6961 CTGCGCTCTTTGAAGGGACATGCACTACCAACATGACTCCCGGAGCTCCCGGAGCTGACTCA 7020
Db 6961 CTGCGCTCTTTGAAGGGACATGCACTACCAACATGACTCCCGGAGCTCCCGGAGCTGACTCA 7020
Qy 7021 TCGAGGCCAACTCTTGTTGGCGGACAGGATGGCGGAAACATCACTCGCTGGAGTCA 7080
Db 7021 TCGAGGCCAACTCTTGTTGGCGGACAGGATGGCGGAAACATCACTCGCTGGAGTCA 7080
Qy 7081 AGAATAAGTAGTAATCTGGAATCTTTGAAACCGCTTCAAGGAGGATGAGAGG 7140
Db 7081 AGAATAAGTAGTAATCTGGAATCTTTGAAACCGCTTCAAGGAGGATGAGAGG 7140
Qy 7141 AGATATCCGTGCGCGGAGATCCCGGAAATCCAGGAAGTTCGCCCTCAGCGTTGCCCA 7200
Db 7141 AGATATCCGTGCGCGGAGATCCCGGAAATCCAGGAAGTTCGCCCTCAGCGTTGCCCA 7200
Qy 7201 TATGGGCACGCGCGGACTACAACTCTTCTAGAGTCCCTGGAAGGACCGGACTACG 7260
Db 7201 TATGGGCACGCGCGGACTACAACTCTTCTAGAGTCCCTGGAAGGACCGGACTACG 7260
Qy 7261 TCCCTCCGGTGTACAGGATGCCATTCGCCACCTTACCAAGGCTCCTCCAATACCACTC 7320
Db 7261 TCCCTCCGGTGTACAGGATGCCATTCGCCACCTTACCAAGGCTCCTCCAATACCACTC 7320
Qy 7321 CACGAGAAAGAGGACGTTGCTGACAGATCCAAATGCTGCTTCTGCTGCGGAGC 7380
Db 7321 CACGAGAAAGAGGACGTTGCTGACAGATCCAAATGCTGCTTCTGCTGCGGAGC 7380
Qy 7381 TCGCCACTAAGACCTTCGGTAGCTCCGGATCGTCCGGCGTTGATAGCGGCACGGGACCG 7440
Db 7381 TCGCCACTAAGACCTTCGGTAGCTCCGGATCGTCCGGCGTTGATAGCGGCACGGGACCG 7440

Qy 7441 CCCTTCTGTGACTTGGCTCCGACGACCGGTGACAAAGGATCCGACGTTGAGTCGTACTCCT 7500
Db 7441 CCCTTCTGTGACTTGGCTCCGACGACCGGTGACAAAGGATCCGACGTTGAGTCGTACTCCT 7500
Qy 7501 CCATGCCCCCTTTGAAAGGGAGCGCGGGGACCCCGATCTCAGCGACGGGTCTTGGTCTA 7560
Db 7501 CCATGCCCCCTTTGAAAGGGAGCGCGGGGACCCCGATCTCAGCGACGGGTCTTGGTCTA 7560
Qy 7561 CCCTGAGTGAGGAGGCTAGTGAGGATGCTGCTGCTGCTCAATGCTCTATACGTGGACAG 7620
Db 7561 CCCTGAGTGAGGAGGCTAGTGAGGATGCTGCTGCTGCTCAATGCTCTATACGTGGACAG 7620
Qy 7621 GCGCCCTGATCAGCCCATGCGCTGCGGAGGAAAGTAAAGCTGCCCATCAACCCGTTGAGCA 7680
Db 7621 GCGCCCTGATCAGCCCATGCGCTGCGGAGGAAAGTAAAGCTGCCCATCAACCCGTTGAGCA 7680
Qy 7681 ACTCTTGTCTGCTGACACAAATGCTCTAOGCCACAAATCCCGCAGCGCAAGCCTCC 7740
Db 7681 ACTCTTGTCTGCTGACACAAATGCTCTAOGCCACAAATCCCGCAGCGCAAGCCTCC 7740
Qy 7741 GGCAGAAAGGTCACTTTTACAGAGTTGCAAGTCCGATGATCAATACGCGAAGCTAC 7800
Db 7741 GGCAGAAAGGTCACTTTTACAGAGTTGCAAGTCCGATGATCAATACGCGAAGCTAC 7800
Qy 7801 TCAAGGAGATGAAGCGGAAAGCGTCCACAGTTAAGGCTTAAGCTTCTATCTATAGAGGAG 7860
Db 7801 TCAAGGAGATGAAGCGGAAAGCGTCCACAGTTAAGGCTTAAGCTTCTATCTATAGAGGAG 7860
Qy 7861 CCTGCAAGTGAACCGCCCACTTTGCGGCCAAATCCAAATTTGGCTATGCGGCAAGGAGC 7920
Db 7861 CCTGCAAGTGAACCGCCCACTTTGCGGCCAAATCCAAATTTGGCTATGCGGCAAGGAGC 7920
Qy 7921 TCCGGAACCTATCCAGCAGGCGGTTAAACACATCCGCTCCGTTGGAGGAGCTTGTGG 7980
Db 7921 TCCGGAACCTATCCAGCAGGCGGTTAAACACATCCGCTCCGTTGGAGGAGCTTGTGG 7980
Qy 7981 AAGACACTGAAACACCAATTTGACACCACTCATGCGCAAAAGTGAAGTCTTCTGCGTCC 8040
Db 7981 AAGACACTGAAACACCAATTTGACACCACTCATGCGCAAAAGTGAAGTCTTCTGCGTCC 8040
Qy 8041 AACACAGAGAGGAGCGCGGACGAGTGGCTTATCGTATTCGCCAGACTTGGAGTTC 8100
Db 8041 AACACAGAGAGGAGCGCGGACGAGTGGCTTATCGTATTCGCCAGACTTGGAGTTC 8100
Qy 8101 GTGTATCGGAGAAATGGCCCTTTACGACGTGGTCTCCACCTTCCCTCAGCGCGTATGG 8160
Db 8101 GTGTATCGGAGAAATGGCCCTTTACGACGTGGTCTCCACCTTCCCTCAGCGCGTATGG 8160
Qy 8161 GCTCCTCATACGGATTTCAATACTTCCCGGAGGAGCGGTCGAGTTCCTGGTGAATACCT 8220
Db 8161 GCTCCTCATACGGATTTCAATACTTCCCGGAGGAGCGGTCGAGTTCCTGGTGAATACCT 8220
Qy 8221 GGAATCAAAAGAAATGCGCTTCTCATATGACACCGCTGTTTGACTCAACGG 8280
Db 8221 GGAATCAAAAGAAATGCGCTTCTCATATGACACCGCTGTTTGACTCAACGG 8280
Qy 8281 TCACCTGAGAGTGAATTCGTGTTGAGGAGTCAATTTACCAATGTTGAGCTTGGCCCCCG 8340
Db 8281 TCACCTGAGAGTGAATTCGTGTTGAGGAGTCAATTTACCAATGTTGAGCTTGGCCCCCG 8340
Qy 8341 AGGCCAGACAGGCCCATTAAGTTCGCTACAGAGGGCTTTATCATGGGGTCCCTGACTA 8400
Db 8341 AGGCCAGACAGGCCCATTAAGTTCGCTACAGAGGGCTTTATCATGGGGTCCCTGACTA 8400
Qy 8401 ACTCAAAAGGGCAGAACTGCGGTTATCGCGGTCGCGGCAAGTGGCGTCTGACGACTA 8460
Db 8401 ACTCAAAAGGGCAGAACTGCGGTTATCGCGGTCGCGGCAAGTGGCGTCTGACGACTA 8460
Qy 8461 GCTCGGTTATACCTCACATGTTTACTTTGAAGGCACTGACAGCTGCGAGCTGCAAGC 8520
Db 8461 GCTCGGTTATACCTCACATGTTTACTTTGAAGGCACTGACAGCTGCGAGCTGCAAGC 8520

Qy	1861	ATTGTTTACCCCAAGCCCTCTGTGTGGGGGCAACCGATCGTTTCGGGTCTCCTACGT	1920
Dd	1861	ATTGTTTACCCCAAGCCCTCTGTGTGGGGGCAACCGATCGTTTCGGGTCTCCTACGT	1920
Qy	1921	ATAGCTGGGGGAGAAATGAGACAGACGTGTGCTCTCAACAACACGCGTCGGCACAAAG	1980
Dd	1921	ATAGCTGGGGGAGAAATGAGACAGACATGATGCTCTCAACAACACGCGTCGGCACAAAG	1980
Qy	1981	GCACCTGTTGGCTGTACATGGATGATAGTACTGGGTTCACCTAAGACGTCGGAGGTC	2040
Dd	1981	GCACCTGTTGGCTGTACATGGATGAATAGTACTGGGTTCACCTAAGACGTCGGAGGTC	2040
Qy	2041	CCCCGTGTAACTCCGGGGGGTCCGTAAACCGCACTTGATCTGCCGCCACGACTGCTCC	2100
Dd	2041	CCCCGTGTAACTCCGGGGGGTCCGTAAACCGCACTTGATCTGCCGCCACGACTGCTCC	2100
Qy	2101	GGAAGCACCCCGAGGCTACTTACACAATAATGTGGCTCGGGGCCCTGGTGAACCTAGGT	2160
Dd	2101	GGAAGCACCCCGAGGCTACTTACACAATAATGTGGCTCGGGGCCCTGGTGAACCTAGGT	2160
Qy	2161	GCCTAGTAGACTACCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTCCATCT	2220
Dd	2161	GCCTAGTAGACTACCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTCCATCT	2220
Qy	2221	TTAAGGTTAGGATGTATGTGGGGGGTGGAGCACAGGCTCAATGCGCATGCAATGGA	2280
Dd	2221	TTAAGGTTAGGATGTATGTGGGGGGTGGAGCACAGGCTCAATGCGCATGCAATGGA	2280
Qy	2281	CTCAGAGAGCGCTCTAACTTGGAGGACAGGATAGGTAGCACTACCGCTGCTGC	2340
Dd	2281	CTCAGAGAGCGCTCTAACTTGGAGGACAGGATAGGTAGCACTACCGCTGCTGC	2340
Qy	2341	TGCTTACACAGAGTGCGAGTACTGCTGCTTACACAGCTACCGCTTATPCCA	2400
Dd	2341	TGCTTACACAGAGTGCGAGTACTGCTGCTTACACAGCTACCGCTTATPCCA	2400
Qy	2401	CTGGTTTGTATCCATCCATGAGAACTGCTGACGTGCAATACCTGTACGGGTAGGGT	2460
Dd	2401	CTGGTTTGTATCCATCCATGAGAACTGCTGACGTGCAATACCTGTACGGGTAGGGT	2460
Qy	2461	CAGCGTTTGTCTCCTTTGCAATCAAAATGGAGTACATCTGTTGCTTCTCTCTG	2520
Dd	2461	CAGCGTTTGTCTCCTTTGCAATCAAAATGGAGTACATCTGTTGCTTCTCTCTG	2520
Qy	2521	CAGACGCGCGGTCTGCTGCTGCTTGGATGCTGCTGATAGCCAGGCTGAGGCG	2580
Dd	2521	CAGACGCGCGGTCTGCTGCTGCTTGGATGCTGCTGATAGCCAGGCTGAGGCG	2580
Qy	2581	CCTTAGAGAACTTGGTGTCTCAATGCGGCTCGTGGCCGAGCGCATGTTCTCT	2640
Dd	2581	CCTTAGAGAACTTGGTGTCTCAATGCGGCTCGTGGCCGAGCGCATGTTCTCT	2640
Qy	2641	CCTTTCTTGTGTTCTTCTGCGCGCTGGTACATTAAGGCGAGCTGCTCTCGGGCG	2700
Dd	2641	CCTTTCTTGTGTTCTTCTGCGCGCTGGTACATTAAGGCGAGCTGCTCTCGGGCG	2700
Qy	2701	CGTATGCTTTTATGGCGTATGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2760
Dd	2701	CGTATGCTTTTATGGCGTATGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2760
Qy	2761	CTTAGCGCTTGGACCGGAGATGGTGTGATCGTGGGGGTGCGGTCTTGTAGTCTGG	2820
Dd	2761	CTTAGCGCTTGGACCGGAGATGGTGTGATCGTGGGGGTGCGGTCTTGTAGTCTGG	2820
Qy	2821	TATCTGTACCTTGTACCATACTAATAAGTGTCTCTCTCTCTCTCTCTCTCTCTCT	2880
Dd	2821	TATCTGTACCTTGTACCATACTAATAAGTGTCTCTCTCTCTCTCTCTCTCTCTCT	2880
Qy	2881	AATACTTTTATACAGAGCGGAGCGCATGCAAGTGTGGGTGCGGCTTCCCTTCAAGTTC	2940
Dd	2881	AATACTTTTATACAGAGCGGAGCGCATGCAAGTGTGGGTGCGGCTTCCCTTCAAGTTC	2940
Qy	2941	GGGAGGCGCGGATGCCATCATCTCTCAGCTGTGGGTTCATCCAGAGTTAATTTTG	3000
Dd	2941	GGGAGGCGCGGATGCCATCATCTCTCAGCTGTGGGTTCATCCAGAGTTAATTTTG	3000
Qy	3001	ACATCACAAACTCTCTCTCGCATACTCGGCCCGCTCATGTGCTCAGGCTGGCATAA	3060
Dd	3001	ACATCACAAACTCTCTCTCGCATACTCGGCCCGCTCATGTGCTCAGGCTGGCATAA	3060
Qy	3061	CGAGAGTCCGCTACTTCTGTGGCGCTCAAGGGCTCATTCATGAAGCTGGCGCTGACAGTA	3120
Dd	3061	CGAGAGTCCGCTACTTCTGTGGCGCTCAAGGGCTCATTCATGAAGCTGGCGCTGACAGTA	3120
Qy	3121	AAAGTCGCGGGGGTCAATTATGTCAAATGTCTTCATGAAGCTGGCGCTGACAGTA	3180
Dd	3121	AAAGTCGCGGGGGTCAATTATGTCAAATGTCTTCATGAAGCTGGCGCTGACAGTA	3180
Qy	3181	CGTAGCTTTATAACCATCTTACCCCACTGCGGACATGTTGGGTCTACCCGTCTCCGCGGAGGG	3240
Dd	3181	CGTAGCTTTATAACCATCTTACCCCACTGCGGACATGTTGGGTCTACCCGTCTCCGCGGAGGG	3240
Qy	3241	TTGCGGTGGCGGTAGAGCCCGTCTCTTCCGCCATGAGACCAAGGTCATCACTGGG	3300
Dd	3241	TTGCGGTGGCGGTAGAGCCCGTCTCTTCCGCCATGAGACCAAGGTCATCACTGGG	3300
Qy	3301	GAGCAGACACCGCTGCTGTGGGACATCATCTTGGGTCTACCCGTCTCCGCGGAGGG	3360
Dd	3301	GAGCAGACACCGCTGCTGTGGGACATCATCTTGGGTCTACCCGTCTCCGCGGAGGG	3360
Qy	3361	GGAGGAGATATTTTGGGACCGGCTGATGTCCTGAAGGCAAGGTCGACCTCTCTG	3420
Dd	3361	GGAGGAGATATTTTGGGACCGGCTGATGTCCTGAAGGCAAGGTCGACCTCTCTG	3420
Qy	3421	CGGCCATCAGGCTTACTCCCAACCGCGGGGTACTTGGTTGCACTCATCTAGCC	3480
Dd	3421	CGGCCATCAGGCTTACTCCCAACCGCGGGGTACTTGGTTGCACTCATCTAGCC	3480
Qy	3481	TCACAGCGCGGACAAAGAACCCAGGTCGAAGGGGAGGTTCAAGTGGTTTCTACCGCAACAC	3540
Dd	3481	TCACAGCGCGGACAAAGAACCCAGGTCGAAGGGGAGGTTCAAGTGGTTTCTACCGCAACAC	3540
Qy	3541	AAATCTTCTTGGCGAAGGCTGATCAACCGGCTGCTGGAGTCTACCATGGCGCTGGT	3600
Dd	3541	AAATCTTCTTGGCGAAGGCTGATCAACCGGCTGCTGGAGTCTACCATGGCGCTGGT	3600
Qy	3601	CGAGACCTTAGCGGTCCAAAAGGTCCTCAATCACCACAAATGTACACCAATGTAGACCTGG	3660
Dd	3601	CGAGACCTTAGCGGTCCAAAAGGTCCTCAATCACCACAAATGTACACCAATGTAGACCTGG	3660
Qy	3661	ACCTCGTGGCTGGCAGGCGCGCCCGCGGCGCGCTCCATGACACCAATGAGCTGGCA	3720
Dd	3661	ACCTCGTGGCTGGCAGGCGCGCCCGCGGCGCGCTCCATGACACCAATGAGCTGGCA	3720
Qy	3721	GCTCGGACCTTTTACTTGGTCAAGACATGCTGATGTCATTCGGTTCGGCGGAGCGG	3780
Dd	3721	GCTCGGACCTTTTACTTGGTCAAGACATGCTGATGTCATTCGGTTCGGCGGAGCGG	3780
Qy	3781	ACACGAGGGAAGTCTACTCTCCCGAGGCGCGCTCTCTACCTGAAAGGTCCTCTGGGTG	3840
Dd	3781	ACACGAGGGAAGTCTACTCTCCCGAGGCGCGCTCTCTACCTGAAAGGTCCTCTGGGTG	3840
Qy	3841	GTCCATTTGCTTTCCTTTCGGGCGACGTCGTTGGCGGTCTTCCGGGTGCTGTGTGCAACC	3900
Dd	3841	GTCCATTTGCTTTCCTTTCGGGCGACGTCGTTGGCGGTCTTCCGGGTGCTGTGTGCAACC	3900
Qy	3901	GGGGGTTCGAAAGCGGTGACATTCATCCCTGTGAGTCTATCGAAACTACCATGCGGT	3960
Dd	3901	GGGGGTTCGAAAGCGGTGACATTCATCCCTGTGAGTCTATCGAAACTACCATGCGGT	3960
Qy	3961	CTCCGCTCTTACACAGAACTCACCACCGCGGTGTACCGCAGACATTCACAGTGGCAC	4020
Dd	3961	CTCCGCTCTTACACAGAACTCACCACCGCGGTGTACCGCAGACATTCACAGTGGCAC	4020
Qy	4021	ATCTGCACTCTCTTCTGCGCAGCGCAAGAGCAACAAAGTGGCGGTGGGTATGACGCC	4080

QY 6241 ACTGCTCTACGCGATGCTCCGGCTGCTGCTAAGGATGTTGGGATTCGGATATGACGG 6300
 Db 6241 ACTGCTCTACGCGATGCTCCGGCTGCTGCTAAGGATGTTGGGATTCGGATATGACGG 6300
 QY 6301 TGTGAGCTGACTCAAGACTGCTCCAGTCCAACTCCTCGCGGGGTACCGGGAGTCC 6360
 Db 6301 TGTGAGCTGACTCAAGACTGCTCCAGTCCAACTCCTCGCGGGGTACCGGGAGTCC 6360
 QY 6361 CTCTTCTCTCATGCCAACGCGGTACAGGAGTCTGGGGGGGACGGCATCATCAAA 6420
 Db 6361 CTCTTCTCTCATGCCAACGCGGTACAGGAGTCTGGGGGGGACGGCATCATCAAA 6420
 QY 6421 CACCTGCGCATCGGAGACACATCGCGGACATGTCAAAGAGTTCCTCAGGATCG 6480
 Db 6421 CACCTGCGCATCGGAGACACATCGCGGACATGTCAAAGAGTTCCTCAGGATCG 6480
 QY 6481 TAGGGCTAGAACTGCAGCAACACGCTGGGACGGAAGTTCGCCATCAACGCAATACCA 6540
 Db 6481 TAGGGCTAGAACTGCAGCAACACGCTGGGACGGAAGTTCGCCATCAACGCAATACCA 6540
 QY 6541 CGGAGCTTGCAACCTCCCGCGCGCCAACTATTCCAGGGGCTATGGCGGGTGGCTG 6600
 Db 6541 CGGAGCTTGCAACCTCCCGCGCGCCAACTATTCCAGGGGCTATGGCGGGTGGCTG 6600
 QY 6601 CTGAGGAGTACGAGGTTACGGCTGCGGGGATTCCTCACTAGTGCACGGGCAATGACCA 6660
 Db 6601 CTGAGGAGTACGAGGTTACGGCTGCGGGGATTCCTCACTAGTGCACGGGCAATGACCA 6660
 QY 6661 CTGACACGTTAAGTGGCCATGCGAGTTCGGCGCCCGGAATTCCTCACTAGGAGTGGATG 6720
 Db 6661 CTGACACGTTAAGTGGCCATGCGAGTTCGGCGCCCGGAATTCCTCACTAGGAGTGGATG 6720
 QY 6721 GAGTGCCTTGCAACGTTACGCTCGCGGCTGCGCAACCTCTTCTACGGGAGACGTTACCT 6780
 Db 6721 GAGTGCCTTGCAACGTTACGCTCGCGGCTGCGCAACCTCTTCTACGGGAGACGTTACCT 6780
 QY 6781 TCCAGGTCGGCTCAACCTACTGCTGCGGCTGCGCAACCTCTTCTACGGGAGACGTTACCT 6840
 Db 6781 TCCAGGTCGGCTCAACCTACTGCTGCGGCTGCGCAACCTCTTCTACGGGAGACGTTACCT 6840
 QY 6841 ACCTAAACAGTCTTACTTCCATGCTACCGATTCCTCCATTCACATACAGCAGACGGCTA 6900
 Db 6841 ACCTAAACAGTCTTACTTCCATGCTACCGATTCCTCCATTCACATACAGCAGACGGCTA 6900
 QY 6901 AGCGTAGGCTGGTAGAGGTTCTCCCGCTCTTTAGCCAGCTCATACCTAGCAGGCTTGT 6960
 Db 6901 AGCGTAGGCTGGTAGAGGTTCTCCCGCTCTTTAGCCAGCTCATACCTAGCAGGCTTGT 6960
 QY 6961 CTGCGCTTCTTTGAGGCGACATGCACTACCCACATGATCCCGGACGCTGACCTCA 7020
 Db 6961 CTGCGCTTCTTTGAGGCGACATGCACTACCCACATGATCCCGGACGCTGACCTCA 7020
 QY 7021 TCGAGGCCAACCTTTGCGGGCAGAGATGGCGGAAACATCTCGGTGGAGTCA 7080
 Db 7021 TCGAGGCCAACCTTTGCGGGCAGAGATGGCGGAAACATCTCGGTGGAGTCA 7080
 QY 7081 AGAATAAGGTAGTATTCTGACTCTTTGAAACGCTTCACGCGAGGGGATGAGAGG 7140
 Db 7081 AGAATAAGGTAGTATTCTGACTCTTTGAAACGCTTCACGCGAGGGGATGAGAGG 7140
 QY 7141 AGATATCCCTCGCGGGAGATCTTCGGAATAATCAGGAAGTTCCTCCTCAGCGTTGCCCA 7200
 Db 7141 AGATATCCCTCGCGGGAGATCTTCGGAATAATCAGGAAGTTCCTCCTCAGCGTTGCCCA 7200
 QY 7201 TATGGGACGCGCGGACTCAATCTCTCCACTGCTAGAGTCTTGGGAAGACCGGACTAG 7260
 Db 7201 TATGGGACGCGCGGACTCAATCTCTCCACTGCTAGAGTCTTGGGAAGACCGGACTAG 7260
 QY 7261 TCCCTCCGTTGTAACGAGATGCCATTCGACCTACAGGCTCTCCAAATACACACCTC 7320
 Db 7261 TCCCTCCGTTGTAACGAGATGCCATTCGACCTACAGGCTCTCCAAATACACACCTC 7320

QY 7321 CACGAGAAAAGAGACGCTTGTCTCTGACAGAAATCCAAATGTGTCTCTGCTTGGCGAGC 7380
 Db 7321 CACGAGAAAAGAGACGCTTGTCTCTGACAGAAATCCAAATGTGTCTCTGCTTGGCGAGC 7380
 QY 7381 TCGCCACTAAGACCTTCGCTAGCTCCGATCGTCCGCGCTTGTATAGCGGACGCGGACCG 7440
 Db 7381 TCGCCACTAAGACCTTCGCTAGCTCCGATCGTCCGCGCTTGTATAGCGGACGCGGACCG 7440
 QY 7441 CCTTCTCTGACCTGCTCCGACGAGTGAAGAAAGATTCGACGCTTGTAGTCTGACTCTCT 7500
 Db 7441 CCTTCTCTGACCTGCTCCGACGAGTGAAGAAAGATTCGACGCTTGTAGTCTGACTCTCT 7500
 QY 7501 CCATGCCCCCTTGAAGGGAGCGGGGACCCGATCTCAGCGAGCGCTCTTGGTCTTA 7560
 Db 7501 CCATGCCCCCTTGAAGGGAGCGGGGACCCGATCTCAGCGAGCGCTCTTGGTCTTA 7560
 QY 7561 CGGTAGTACGAGGCTAGTGAAGATGTCTGTGTCTGCTCAATGTCTATACGTGACACAG 7620
 Db 7561 CGGTAGTACGAGGCTAGTGAAGATGTCTGTGTCTGCTCAATGTCTATACGTGACACAG 7620
 QY 7621 GGGCCCTGATCAGCCATCGCTGGGAGGAAAGTAAAGCTGCCATCAACCCGTTGAGCA 7680
 Db 7621 GGGCCCTGATCAGCCATCGCTGGGAGGAAAGTAAAGCTGCCATCAACCCGTTGAGCA 7680
 QY 7681 ACTTCTTGTCTGCTCAGCACACATGCTACGCCAACAATCCCGACGCGAAGCTCTC 7740
 Db 7681 ACTTCTTGTCTGCTCAGCACACATGCTACGCCAACAATCCCGACGCGAAGCTCTC 7740
 QY 7741 GGCAGAAAGAGTCACTTTGACAGATTGCAAGTCTCTGATGATCAATTACCGGAGAGTAC 7800
 Db 7741 GGCAGAAAGAGTCACTTTGACAGATTGCAAGTCTCTGATGATCAATTACCGGAGAGTAC 7800
 QY 7801 TCAAGGAGATGAAGCGAAGGCTCCACAGTTAAGCTAAGCTTCTATCTATAGAGGAGG 7860
 Db 7801 TCAAGGAGATGAAGCGAAGGCTCCACAGTTAAGCTTCTATCTATAGAGGAGG 7860
 QY 7861 CTTGCAAGCTGACGCGCCCACTTCGCGCCAAATTCCAAATTTGGCTATGGGCGAAAGAGC 7920
 Db 7861 CTTGCAAGCTGACGCGCCCACTTCGCGCCAAATTCCAAATTTGGCTATGGGCGAAAGAGC 7920
 QY 7921 TCCGGAACCTTATCCAGCAGGCGGCTTAAACACATCCGCTCCGCTGGGAGGAGTCTGCTGG 7980
 Db 7921 TCCGGAACCTTATCCAGCAGGCGGCTTAAACACATCCGCTCCGCTGGGAGGAGTCTGCTGG 7980
 QY 7981 AAGACACTGAAAACCAATTAACACCAATTCACCAATTCGCAAAAGTGAAGTCTTCTGCGCTCC 8040
 Db 7981 AAGACACTGAAAACCAATTAACACCAATTCACCAATTCGCAAAAGTGAAGTCTTCTGCGCTCC 8040
 QY 8041 AACAGAGAGAGGAGCGCGCAAGCAGCTCGCTTATCGTATTCAGAGACTTGGGAGTTC 8100
 Db 8041 AACAGAGAGAGGAGCGCGCAAGCAGCTCGCTTATCGTATTCAGAGACTTGGGAGTTC 8100
 QY 8101 GTGTATGCGAAGATGCGCTTTTACGACCTGCTTCAACGCTTCTCAACGCTTCTCAGCGCTGATGG 8160
 Db 8101 GTGTATGCGAAGATGCGCTTTTACGACCTGCTTCAACGCTTCTCAGCGCTGATGG 8160
 QY 8161 GCTCCTCTATACGAGATTCATACCTTCCCGCAGCAGCGCTCGAGTCTCTGCTGATTAACCT 8220
 Db 8161 GCTCCTCTATACGAGATTCATACCTTCCCGCAGCAGCGCTCGAGTCTCTGCTGATTAACCT 8220
 QY 8221 GGAATCAAGAAATGCGCTTATGGCTTCTCATATGACACCGCTGTTTGGACTCAACGG 8280
 Db 8221 GGAATCAAGAAATGCGCTTATGGCTTCTCATATGACACCGCTGTTTGGACTCAACGG 8280
 QY 8281 TCACCTGAGAGTCAATTCGCTTGTGAGGAGTCAATTTACCAATGTGTGACTTGGCCCCCG 8340
 Db 8281 TCACCTGAGAGTCAATTCGCTTGTGAGGAGTCAATTTACCAATGTGTGACTTGGCCCCCG 8340
 QY 8341 AGGCCAGACAGGCCATTAAGTCTGCTCAAGAGCGCTTTACATCGGCGGCTCCCTGACTA 8400
 Db 8341 AGGCCAGACAGGCCATTAAGTCTGCTCAAGAGCGCTTTACATCGGCGGCTCCCTGACTA 8400
 QY 8401 ACTCAAAAGGCGAAGACTGCGGTTATCCCGGTCGCCGCAAGTGGCGTGTGTCAGGACTA 8460

Db 8401 ACTGAAAGGCAAACTGCGGTTATCGCGGTGCGGCGAAGTGGGTGCTGACGACTA 8450
|
Qy 8461 GCTGGGTAATACCCCTACATGTTACTTGAAGGCACATGAGGCTGCTGAGCTGCAAGC 8520
|
Db 8461 GCTGGGTAATACCCCTACATGTTACTTGAAGGCACATGAGGCTGCTGAGCTGCAAGC 8520
|
Qy 8521 TCCAGGACTGACGATGCTCGTGAAGGAGAGACCTTGTGTTATCTGTGAAAGCGCG 8580
|
Db 8521 TCCAGGACTGACGATGCTCGTGAAGGAGAGACCTTGTGTTATCTGTGAAAGCGCG 8580
|
Qy 8581 GAACCCAGGAGGATCGGCGGCTTACGAGCCCTTACGAGGCTATGACTAGGATTCG 8640
|
Db 8581 GAACCCAGGAGGATCGGCGGCTTACGAGCCCTTACGAGGCTATGACTAGGATTCG 8640
|
Qy 8641 CCCCCCGGGGATCGGCGGCTTACGAGCCCTTACGAGGCTATGACTAGGATTCG 8700
|
Db 8641 CCCCCCGGGGATCGGCGGCTTACGAGCCCTTACGAGGCTATGACTAGGATTCG 8700
|
Qy 8701 CCAATGTGTCACTCGCGGACGATGATCTGGCAAAAGGATTAATACCTCACCGTGACC 8760
|
Db 8701 CCAATGTGTCACTCGCGGACGATGATCTGGCAAAAGGATTAATACCTCACCGTGACC 8760
|
Qy 8761 CCACCAACCCCTTGCACCGGCTGCGTGGGAGACAGCTAGACACACTCCAGTCAACTCT 8820
|
Db 8761 CCACCAACCCCTTGCACCGGCTGCGTGGGAGACAGCTAGACACACTCCAGTCAACTCT 8820
|
Qy 8821 GCGTAGGCAATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8880
|
Db 8821 GCGTAGGCAATATCATGATGATGATGATGATGATGATGATGATGATGATGATG 8880
|
Qy 8881 ACTTTTCTCCATCTCTAGCTCAAGAGCAACTTGAAGGAGGCTTGGATGTCAGATCT 8940
|
Db 8881 ACTTTTCTCCATCTCTAGCTCAAGAGCAACTTGAAGGAGGCTTGGATGTCAGATCT 8940
|
Qy 8941 ACGGGCTTGCTACTCATGAGCCTTGGAGCCTTGGAGCCTTGGAGCCTTGGAGCCTTGG 9000
|
Db 8941 ACGGGCTTGCTACTCATGAGCCTTGGAGCCTTGGAGCCTTGGAGCCTTGGAGCCTTGG 9000
|
Qy 9001 GTCTTAGCGATTACACTCCAGATGATCTCCAGTGAAGTCAATGAGGCTGCTCAT 9060
|
Db 9001 GTCTTAGCGATTACACTCCAGATGATCTCCAGTGAAGTCAATGAGGCTGCTCAT 9060
|
Qy 9061 GCCTCAGAAACTTGGGATACCACTTGGAGCCTTGGAGCCTTGGAGCCTTGGAGCCTTGG 9120
|
Db 9061 GCCTCAGAAACTTGGGATACCACTTGGAGCCTTGGAGCCTTGGAGCCTTGGAGCCTTGG 9120
|
Qy 9121 CGCTAGGCTACTGCTCCAGGGGGAGGCGGCTTGGAGCCTTGGAGCCTTGGAGCCTTGG 9180
|
Db 9121 CGCTAGGCTACTGCTCCAGGGGGAGGCGGCTTGGAGCCTTGGAGCCTTGGAGCCTTGG 9180
|
Qy 9181 GGGCAGTAAGGACCAAGCTTAACTCACTCCAACTCCCGGCGGCTTGGAGCCTTGGAGCCTTGG 9240
|
Db 9181 GGGCAGTAAGGACCAAGCTTAACTCACTCCAACTCCCGGCGGCTTGGAGCCTTGGAGCCTTGG 9240
|
Qy 9241 CTGGCTGCTTCTGCTGCTTACAGGGGGAGGCGGCTTGGAGCCTTGGAGCCTTGGAGCCTTGG 9300
|
Db 9241 CTGGCTGCTTCTGCTGCTTACAGGGGGAGGCGGCTTGGAGCCTTGGAGCCTTGGAGCCTTGG 9300
|
Qy 9301 GACCCCGCTGCTTCCGTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 9360
|
Db 9301 GACCCCGCTGCTTCCGTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 9360
|
Qy 9361 TCCCAACCGGATGACGGGAGCTAACCACTCCAGGCTTAAAGCCATTTCTTCTTCTTCTTCTTCT 9420
|
Db 9361 TCCCAACCGGATGACGGGAGCTAACCACTCCAGGCTTAAAGCCATTTCTTCTTCTTCTTCTTCT 9420
|
Qy 9421 TTTTCT 9480
|
Db 9421 TTTTCT 9480
|
Qy 9481 TTTTCT 9540
|

Db 9481 TTTTCT 9540
|
Qy 9541 CCGTGAGCCGATGCTGAGAGTGTCTGATCTGCGCTTCTTCTGAGATCACTGT 9595
|
Db 9541 CCGTGAGCCGATGCTGAGAGTGTCTGATCTGCGCTTCTTCTGAGATCACTGT 9595
|
RESULT 4
AF054249 9596 bp RNA linear VRL 02-MAY-1998
LOCUS Hepatitis C virus HC-J4, pCV-J4L4S, complete genome.
DEFINITION
AF054249
ACCESSION
AF054249.1 GI:3098636
VERSION
AF054249.1
KEYWORDS
Hepatitis C virus
SOURCE
Hepatitis C virus
ORGANISM
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 9596)
Yanagi,M., St Claire,M., Shapiro,M., Emerson,S.U., Purcell,R.H. and
Bukh,J.
Transcripts of a chimeric cDNA clone of hepatitis C virus genotype
1b are infectious in vivo
Virology 244 (1), 161-172 (1998)
MEDLINE
98240944
PUBMED
9581788
REFERENCE
2 (bases 1 to 9596)
Yanagi,M. and Bukh,J.
Direct Submission
TITLE
Submitted (18-MAR-1998) Hepatitis Viruses Section, LID, NIAID, NIH,
7 Center Dr., Building 7, Room 201, Bethesda, MD 20892, USA
FEATURES
source
1..9596
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/strain="HC-J4"
/isolate="pCV-J4L4S"
/db_xref="taxon:11103"
/note="non-infectious full-length cDNA"
343..9375
/codon_start=1
/product="polyprotein"
/protein_id="AA15724.1"
/db_xref="GI:3098637"
/translation="MSTNPKPQKTKBNTR3PDQVPPGGQIVGGVLLPRGRPL
GVATKTSRSQPRGRQPIKARPEGRMAQPGYPMPLXGNEGLWAGMLLSPRG
SSPWTDPRRSRNLKVIDTLCTGFPADLMGYIPLVAGPLGGARALAHGVRLLED
GVNYATGNLPGGSPSIFLLALLSLCTIPASAVEVRNVSIGIYHVDLILVTAAPCSAY
VIMHTGCPVPCVRGSSRCWALPTLAARDASVETTTIRHVDLILVTAAPCSAY
VGLDGSIFLNSQLFTSPRRHETVQDCNSIYPGHVSRMAWMMMSPTALVY
SOLLEIQAQVDMVAGHWGLAGLAVYMWGNWAKVLTALLFAGVDETHTTRV
GRTISFTSLPSSGASQKIQLVNTGSHNRTALNCNDSLOTGFLAALFYTHKNSP
GCPERNASCRPIDWFAQMGPTIYEPNSSDQRPYCMHYAPRPGTVPASQVCGPVY
FTSPVVGTTDSRGVPTYSWGENETVDMLNNTTRPPQGNWFGCTWMSGTRKTCGG
PCNIGVGNRTLICPTDCKRHEATYTCGSGPWLTPRCLVDIYRLLWHPCTLNF
SIFKMYGVGVHRLNAACNWTGRCLNEDRSELSFLLSTTEWOLLPALPIT
SIFKMYGVGVHRLNAACNWTGRCLNEDRSELSFLLSTTEWOLLPALPIT
IAQAEALENLAVLNAASVAGAGHILSLFVFFCAAWIKRGLAPGAAYFVGVPLLL
LLALPFRVALDREMAASCGGAVLVLPLTSLPVYKVLTRLIWLVOYFVIRAEAH
MOWVPLNVRGGRDAIILITCAVHPELFDITKLLIALLGLMLVLAQGITRVPYVR
AQGLIRACMLVRKVGSHVYQVAFMKLGAITGVYVYNHUTPLRDNHAGLRDLAVE
PVYSAMETKVTWAGDTACGDIILGLPVSARRKEIFLGPADSLEGQGWRLALPIT
AYSOQRTGVLGCVTISITGRDKQVGEVOVSTATQSFATCINGVCTWTVHAGSK
TLGAKGPTQMTNTVDLIDVQAPPGASMTSCGSDLYLTVRHADVIIVRSRG
DSRGLSLSPRPVITLKGSGGPLLCRSHVGVVFRANVCTRGVAKVADRIPVPSMETT
MRSPVFTDSSPPAVPTQVLAHHAFTGSKSTKVPFAAYAAQGVKVLVLPNSAATL
GFGAYMSKHGIDPNITGVTITGSSITYTGYKFLADGCGSGGAYDILIIICDECHS
TDSITLIGITGLVLDQATAGARLVLAATATPGSVTVPHENIEIIGLSNNGEIPFYK
AIPIEATKGGHLLIFCHSKKKCDLAKLTGLGLNAVAYYRGLDVSIVIPITGVVVA
TDALMTGTFDPSVIDCNTCVTQVDPISLDPITITTTVPQDVSRSORRGTRG
RSGLYRVTGERPFGMFSSVLCEDACAMHETPAETSVRLAYLATPGLPVQC
DHLEFWSVETGLTHDAFLSQTQKQAGDNFFILVATQATVCAQAQAPPSWQWKC
LIRLKPTLHGPTPLLYRLGAVQNEVILTHPTIKYIMACMSADLEIVTSTWLVGGVLA

ALAA YCU TUGSVI VIGRI IILSCKPAVDPDRVLVQEDPMEBCASQLPVI BQGCQLAS
 OFKALGLLITATKQCEAAAPVPSKWRALTEITWAKHMWNPISGIVLAGLITLPGN
 PATAIIMAFITSTIPUTQNTLLINLWNAQAAPPASAPVAGLAVAGAAVGS
 GLRGVLVDLGVAGVAGALVAPKWSGEVSTPDLNLLPAILSPGALVGVVCAAL
 ILRHVGEIGAGVQOMNRLLIAFASRGHNVSVTHVPEDSAAARVUTLPSITITQLLK
 RLFIOWINEDGTPCSCGWLNRVDWMI CVTLDITKLTOSKLLPRLGVFVLPS CQSGY
 GWNRGDQMNTCPGCAIAGHVKNQSNRI VGPRTCSNTEGTPVINA VTYGTPCSY
 APNTSRGALNRVAEYVEVTRGVGVHYVIMDNVCKQ CQVPAEPTFEDVGRVLRH
 YAPACRPLPDLTFOVGLNQLVGSQPCPEBPVDVLITSMLTDPESH TASTAKRRL
 YARGPPLISSASQALSPKATCTHDSQEPDAILBANLRWOMGENTIRVSEEN
 KVVILDSGPELHAGDDEEISVAAEILKSRKFPALPIWARPDLNPLLESSEWSDPY
 VVPPVGLPGLPTKAPLPIPPRRKTEITVLTSTNSVSLAEALIAKTFGSSGSAVGG
 ATAPLADLSQDGGKSDVESHSMPLPEGECDPLDPSGWSVTESEASEDVCCSMS
 YTWTCALTPCAEESKLPIPLNSLRHNHNVYATTSRSASLKQVKVPTDRLQVLO
 DHYRDLKEMAKASVTKVGLSITBEACKLTPPHKASGFGYGAOKDRNLSRAYNHI
 RSVMEDLDETSTPIDTITIMAKSEVFCVQEPKGRKPLVLPDLGVKTSRGMALD
 VVSTLPPQWAGSSTGYSPQSPKRVFNTLWTKSRCKPMFGSTPDRFSTVSTSDIRV
 BESTVQCCLDAEPARQSTILERYIIGELPINXGONGYCRBAGVTSVCSGNTL
 TCYVLTAAACBAKLODCTMLVNGDVLVYICBSAQOEDAAALRAFTEAMTRYSCAPP
 DPQOEDYLELITSCSSNVSVADHASGEVYVLTDPDTPPLTAAKMETARHPINSLH
 GNIINMAYTPLWARMNLWTFHSILIAKQDEKALDCQIYGACVSTIEPLDLPQILRSLH
 GLSAFTLHSGYENLRNVALSCIKLVGPPLTRWHRASRVAKILSSQGGRAATCGRYL
 FNWAVRTKLKUTPIPAASQDLDSGWFVAGYSGCDITVHLSRARPRFWPTPCLLHLSVGV
 TIVYLLPNR

ORIGIN

Query Match	98.7%;	Score 9473.6;	DB 14;	Length 9596;	
Best local similarity	99.3%;	Pred. No. 0;			
Matches 9526;	Conservative	0;	Mismatches 69;	Indels 1;	Gaps 1;
QY	1	GCACGCCCCCTGATGGGGCGCACACTCCACATCAATCACTCCCTCTGAGGAACACTG	60		
Db	1	GCACGCCCCCTGATGGGGCGGACACTCCACATCAATCACTCCCTCTGAGGAACACTG	60		
QY	61	TCCTACGCGAAGAAGCGCTTAGCCATGGGCTAGTATGAGTGTGTCAGCCTCCAGGAC	120		
Db	61	TCCTACGCGAAGAAGCGCTTAGCCATGGGCTAGTATGAGTGTGTCAGCCTCCAGGAC	120		
QY	121	CCCCCTCCCGGGAGAGCCATAGTGGTCTCGGAAACCGGTGAGTACACCGGAATTGCCAG	180		
Db	121	CCCCCTCCCGGGAGAGCCATAGTGGTCTCGGAAACCGGTGAGTACACCGGAATTGCCAG	180		
QY	181	GACGACCGGGTCCTTTCTTTGGATC- AACCCGCTCAATGCCCTGGAGATTGGCGGTGCCCC	239		
Db	181	GACGACCGGGTCCTTTCTTTGGATCAAACCCGCTCAATGCCCTGGAGATTGGCGGTGCCCC	240		
QY	240	CGGAGACTGCTAGCCGAGTACTGTGGTTCGCGAAGAGCCCTTGGTACTGCTCGCTGATAG	299		
Db	241	CGGAGACTGCTAGCCGAGTACTGTGGGTTCGCGAAGAGCCCTTGGTACTGCTCGCTGATAG	300		
QY	300	GGTGTCTTCGAGTGCCTCCGGGAGGTCTCGTAGACCGTGCACCATGAGCAGGAATCCTAAA	359		
Db	301	GGTGTCTTCGAGTGCCTCCGGGAGGTCTCGTAGACCGTGCACCATGAGCAGGAATCCTAAA	360		
QY	360	CCTCAAGAAAAAACAACAGTTAAACCAACCGCCGCCACAGGACGTCAAGTTCCTCCGGC	419		
Db	361	CCTCAAGAAAAAACAACAGTTAAACCAACCGCCGCCACAGGACGTCAAGTTCCTCCGGC	420		
QY	420	GGTGTCTCAGATCGTTGGTGGAGTTTACCTGTTCGCGCGCAGGGGCCACAGGTGTGGTGTG	479		
Db	421	GGTGTCTCAATTCCTTGGTGGAGTTTACCTGTTCGCGCGCAGGGGCCACAGGTGTGGTGTG	480		
QY	480	CGCGGACTAGGAAGGCTTCCGAGCGGTGCGAACTTCGTGGAAGCGGCAACACTATCCCA	539		
Db	481	CGCGGACTAGGAAGACTTCCGAGCGGTGCGAACTTCGTGGAAGCGGCAACACTATCCCA	540		
QY	540	AAGGTCGCGCACCCGAGGCGAGGCGCTGGGCTCAGCCCGGGTACCTTGGCGCCCTCTAT	599		
Db	541	AAGGTCGCGCGACCCGAGGCGAGGCGCTGGGCTCAGCCCGGGTACCTTGGCGCCCTCTAT	600		
QY	600	GGCAATGAGGGGCTGGGTTGGCAGAGTGGCTCCTGTCAACCCCGGGGTCCCGGCGTAGT	659		
Db	601	GGCAATGAGGCTTGGGTTGGCAGAGTGGCTCCTGTCAACCCCGGGGTCCCGGCGTAGT	660		

3901 CGGGGGTGGGAGGCGGTGACTTCAATACCCGTTGAGTCTATGGAACATCAATCGG 3960 Db
3960 TCTCCGGTCTTACAGACAACATCAACCCCGCGGTGTACCGCAGACATTCAGTGGCA 4019 Qy
3961 TCTCCGGTCTTACAGACAACATCAACCCCGCGGTGTACCGCAGACATTCAGTGGCA 4020 Db
4020 CATCTGACGCTCTTCTGGAGCGGCAAGACACAAAGTCCGCGTGCATGACGCC 4079 Qy
4021 CATCTGACGCTCTTCTGGAGCGGCAAGACACAAAGTCCGCGTGCATGACGCC 4080 Db
4080 CAAAGGTACAAAGTGTCTGCTCTTAAATCCGTCGCTGCGCGCCACCTTAGGGTTGGGGCG 4139 Qy
4081 CAAAGGTACAAAGTGTCTGCTCTTAAATCCGTCGCTGCGCGCCACCTTAGGGTTGGGGCG 4140 Db
4140 TATATGTCCAAGGACACACGCTATPCGACCTTAATCAGAACTGGGGTATGAGCAATAC 4199 Qy
4141 TATATGTCCAAGGACACACGCTATPCGACCTTAATCAGAACTGGGGTATGAGCAATAC 4200 Db
4200 ACGGGCGGCTCCATTAGTACTCTCACTATGGCAAGTTCCTTCCGCAAGGTGCTGTCT 4259 Qy
4201 ACGGGCGGCTCCATTAGTACTCTCACTATGGCAAGTTCCTTCCGCAAGGTGCTGTCT 4260 Db
4260 GGGGCGGCTATGACATCAATATGTAAGTGCACCTCAACTGACTCGACTCAATC 4319 Qy
4261 GGGGCGGCTATGACATCAATATGTAAGTGCACCTCAACTGACTCGACTCAATC 4320 Db
4262 TTGGGCATCGCACAGTCTTGACCAAGCGGAGACGGCTGGAGCGCGGCTGCTGCTC 4379 Qy
4321 TTGGGCATCGCACAGTCTTGACCAAGCGGAGACGGCTGGAGCGCGGCTGCTGCTC 4380 Db
4380 GCCACCGGTACACCTCGGGATCGTTACCGTGCACACCCCAATATGAGGAATAGGC 4439 Qy
4381 GCCACCGGTACACCTCGGGATCGTTACCGTGCACACCCCAATATGAGGAATAGGC 4440 Db
4440 CTGTCCCAATATGAGAGATCCCTTCTATGCAAGAACCATCCCATGAGGCCATCAG 4499 Qy
4441 CTGTCCCAATATGAGAGATCCCTTCTATGCAAGAACCATCCCATGAGGCCATCAG 4500 Db
4500 GGGGGGAGGATCTCTATTTCTGCAATTCGAAGAAATATGAGAGCTCGCGCAAG 4559 Qy
4501 GGGGGGAGGATCTCTATTTCTGCAATTCGAAGAAATATGAGAGCTCGCGCAAG 4560 Db
4560 CTGACAGGCTCGGACTGAAAGCTGTAGCATATTAACGGGCTTGAATGTCCCTCAT 4619 Qy
4561 CTGACAGGCTCGGACTGAAAGCTGTAGCATATTAACGGGCTTGAATGTCCCTCAT 4620 Db
4620 CGGCTATCGGAGACGCTGCTGTGCGCAACAGACGCTCTAATACGGGTTTCCGCGC 4679 Qy
4621 CGGCTATCGGAGACGCTGCTGTGCGCAACAGACGCTCTAATACGGGTTTCCGCGC 4680 Db
4680 GATTTGACTCAGTATGATCAATATGATGTCACCAAGCAAGTGTGAGCTCGAGCTTG 4739 Qy
4681 GATTTGACTCAGTATGATCAATATGATGTCACCAAGCAAGTGTGAGCTCGAGCTTG 4740 Db
4740 GATCCACCTTCACTATGAGAGACGACGCTGCGCCCAAGACGGGTGCGGCTCGCAA 4799 Qy
4741 GATCCACCTTCACTATGAGAGACGACGCTGCGCCCAAGACGGGTGCGGCTCGCAA 4800 Db
4800 CGGCGAGGTAGAACTGGCGGGGTAGAGTGGCATATCAGGTTTGTGACTCAGAGAA 4859 Qy
4801 CGGCGAGGTAGAACTGGCGGGGTAGAGTGGCATATCAGGTTTGTGACTCAGAGAA 4860 Db
4860 CGGCGCTCGGGCATGTTTGGATTTCTTGGTCTGTTGAGTGTATGACGGGCTGTGCT 4919 Qy
4861 CGGCGCTCGGGCATGTTTGGATTTCTTGGTCTGTTGAGTGTATGACGGGCTGTGCT 4920 Db
4920 TGGTATGAGCTCAGCGCGCTGAGACCTCGGTTAGGTTGCGGGTTACCTAAATACCA 4979 Qy
4921 TGGCATCAGTCAACCGCGCTGAGACCTCGGTTAGGTTGCGGGTTACCTAAATACCA 4980 Db
4980 GGGTTGCGGCTCTGACAGGACCATCTGGAGTTCTGGAGAGCGTCTTCAAGCGCTCACC 5039 Qy
4981 GGAATGCGGCTCTGCGAGGACCATCTGGAGTTCTGGAGAGCGTCTTCAAGCGCTCACC 5040 Db

5040 CACATAGATGCCACTTCTCTGTCCAGACTAAACAGGACGAGACAACTTCTTTACCTG 5099 Qy
5041 CACATAGATGCCACTTCTCTGTCCAGACTAAACAGGACGAGACAACTTCTTTACCTG 5100 Db
5100 GTGGCATATCAAGTCTAAGTGTGCGCAGGGCTCAAGTCCACCTCCATCGTGGAGCAA 5159 Qy
5101 GTGGCATATCAAGTCTAAGTGTGCGCAGGGCTCAAGTCCACCTCCATCGTGGAGCAA 5160 Db
5160 ATGTGGAAAGTGTCTCATACGGCTGAAACCTTACACTGCAAGGGCCCAACCCCTGCTGAT 5219 Qy
5161 ATGTGGAAAGTGTCTCATACGGCTGAAACCTTACACTGCAAGGGCCCAACCCCTGCTGAT 5220 Db
5220 AGGCTAGAGCGGCTCCAAAATGAGGTCACTTCCACACCCCATTAATACATCATG 5279 Qy
5221 AGGCTAGAGCGGCTCCAAAATGAGGTCACTTCCACACCCCATTAATACATCATG 5280 Db
5280 GCATGCATGTCGCGCTGACCTGAGGTGCTCACTAGCACCTGGTGTGCTGGCGGAGTTC 5339 Qy
5281 GCATGCATGTCGCGCTGACCTGAGGTGCTCACTAGCACCTGGTGTGCTGGCGGAGTTC 5340 Db
5340 CTTGCAGCTTTGGCGGCTATCTGCTGACGACAGGAGTGTGCTTCTTGGGCGAGATC 5400 Qy
5341 CTTGCAGCTTTGGCGGCTATCTGCTGACGACAGGAGTGTGCTTCTTGGGCGAGATC 5400 Db
5400 ATCTTGTCCGGAGAGCCAGCTGCTTCCGACAGGAAAGTCTCTACACAGGAGTTCGAT 5459 Qy
5401 ATCTTGTCCGGAGAGCCAGCTGCTTCCGACAGGAAAGTCTCTACACAGGAGTTCGAT 5460 Db
5460 GAGATGGAGAGTGTGCTTCCACAACTTCTTCAATCGACAGGAAATGACGTTCGCGAG 5519 Qy
5461 GAGATGGAGAGTGTGCTTCCACAACTTCTTCAATCGACAGGAAATGACGTTCGCGAG 5520 Db
5520 CAATTCAGCAAAAGGCGCTCGGCTTGTGCAAAAGGCGCTTCTGGGCGAAGACATGCG 5579 Qy
5521 CAATTCAGCAAAAGGCGCTCGGCTTGTGCAAAAGGCGCTTCTGGGCGAAGACATGCG 5580 Db
5580 GCTCCCTGTTGAGTCCAAAGTGGCGAGCGCTTCTGAGACCTTCTGGGCGAAGACATGCG 5639 Qy
5581 GCTCCCTGTTGAGTCCAAAGTGGCGAGCGCTTCTGAGACCTTCTGGGCGAAGACATGCG 5640 Db
5640 AATTTCTATCAGCGGAAATACAGTCTTCTGAGACCTTCTGGGCGAAGACATGCG 5699 Qy
5641 AATTTCTATCAGCGGAAATACAGTCTTCTGAGACCTTCTGGGCGAAGACATGCG 5700 Db
5700 ATAGCATCATTTGAGTGGCAATTTACAGCTTCTTCACTAGCGGCTTCTGAGAAACCCGCG 5759 Qy
5701 ATAGCATCATTTGAGTGGCAATTTACAGCTTCTTCACTAGCGGCTTCTGAGAAACCCGCG 5760 Db
5760 CTCCTGTTTAAACATCTTGGGGGATGGGTGGCTGCCAACTCGTCTCTCCAGCGCTGCG 5819 Qy
5761 CTCCTGTTTAAACATCTTGGGGGATGGGTGGCTGCCAACTCGTCTCTCCAGCGCTGCG 5820 Db
5820 TCAGCTTTTAAACATCTTGGGGGATGGGTGGCTGCCAACTCGTCTCTCCAGCGCTGCG 5879 Qy
5821 TCAGCTTTTAAACATCTTGGGGGATGGGTGGCTGCCAACTCGTCTCTCCAGCGCTGCG 5880 Db
5880 GTGCTCGTGGACATCTTGGCGGCTTATGGGCGAGGGGTAGCGGCGCTGCTGCGGCTTT 5939 Qy
5881 GTGCTCGTGGACATCTTGGCGGCTTATGGGCGAGGGGTAGCGGCGCTGCTGCGGCTTT 5940 Db
5940 AAGGTCATGAGCGGAGGTGCGCTTCCACCGAGGACCTGGTCACTTCTGCTGCGCATC 5999 Qy
5941 AAGGTCATGAGCGGAGGTGCGCTTCCACCGAGGACCTGGTCACTTCTGCTGCGCATC 6000 Db
6000 CTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6059 Qy
6001 CTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060 Db
6060 GGGCGGAGAGGGGCTGTGAGTGGATGAACCGGCTGATGAGTTCGCTTCTGCGGGGT 6119 Qy
6061 GGGCGGAGAGGGGCTGTGAGTGGATGAACCGGCTGATGAGTTCGCTTCTGCGGGGT 6120 Db

Db	8281	GTACCGAGTGCATTCGTTGTAGGAGTCAATTTACCAATGTTGTGACTTGGCCCC	8340
Qy	8340	GAGCCACAGAGGCCATTAAGTCCGTCCATCAGAGCGCTTTACATCGGGGCTCCCTGACT	8399
Db	8341	GAGCCACAGAGGCCATTAAGTCCGTCCATCAGAGCGCTTTACATCGGGGCTCCCTGACT	8400
Qy	8400	AACTCMAAGGCGAACTGGGTATCCGCGGTGCCGCGCAAGTGGCGTGTGACGACT	8459
Db	8401	AACTCMAAGGCGAACTGGGTATCCGCGGTGCCGCGCAAGTGGCGTGTGACGACT	8460
Qy	8460	AGCTCGGTATACCTCACATGTTACTGTGAAGGCCACTGAGACCTGTGACTGCAAG	8519
Db	8461	AGCTCGGTATACCTCACATGTTACTGTGAAGGCCACTGAGACCTGTGACTGCAAG	8520
Qy	8520	CTCCAGGACTGCACGATGCTCGTGAAGGAGACGACTTGTCTTATCTGTGAAAGCGCG	8579
Db	8521	CTCCAGGACTGCACGATGCTCGTGAAGGAGACGACTTGTCTTATCTGTGAAAGCGCG	8580
Qy	8580	GGAAACCCAGGAGATGGCGGCGCTTACGAGCCTTCCAGGAGCTATGACTAGGTATTC	8639
Db	8581	GGAAACCCAGGAGATGGCGGCGCTTACGAGCCTTCCAGGAGCTATGACTAGGTATTC	8640
Qy	8640	GCCCCCCCCGGGATCCGCCCAACAGAAATAGCACTGGAGCTGATACATCATGTTCC	8699
Db	8641	GCCCCCCCCGGGATCCGCCCAACAGAAATAGCACTGGAGCTGATACATCATGTTCC	8700
Qy	8700	TCCATGTCAGTCCGCGACGATGATCTGGCAAAAGGTATACACTCACCCGTGAC	8759
Db	8701	TCCAAATGTGTCAGTCCGCGACGATGATCTGGCAAAAGGTATACACTCACCCGTGAC	8760
Qy	8760	CCACACACCCCTTGCACGCGGTGGTGGGAGACAGCTAGACACACTCCAAATCACTCT	8819
Db	8761	CCACACACCCCTTGCACGCGGTGGTGGGAGACAGCTAGACACACTCCAAATCACTCT	8820
Qy	8820	TGGCTAGGCATATCATATGTTATGCCCCACCTATGGCAAGGATGATTTCATGACT	8879
Db	8821	TGGCTAGGCATATCATATGTTATGCCCCACCTATGGCAAGGATGATTTCATGACT	8880
Qy	8880	CACCTTTCTCCATCTCTTAGCTCAAGAGCACTTCGAAAAGCCCTGGATGTGCAGATC	8939
Db	8881	CACCTTTCTCCATCTCTTAGCTCAAGAGCACTTCGAAAAGCCCTGGATGTGCAGATC	8940
Qy	8940	TACGGGCTTGCTACTCCATTTAGCCCACTTGACCTACCTCAGATCAATGAAAGCTCAT	8999
Db	8941	TACGGGCTTGCTACTCCATTTAGCCCACTTGACCTACCTCAGATCAATGAAAGCTCAT	9000
Qy	9000	GGCTTAGGCATTTACCTCCACAGTACTCTCCAGTGGAGTCAATAGGGTGGCTTCA	9059
Db	9001	GGCTTAGGCATTTACCTCCACAGTACTCTCCAGTGGAGTCAATAGGGTGGCTTCA	9060
Qy	9060	TGCTTCAGGAACTTGGGTACACCTTGGAACTGGAGACATCGGGCCAGAGTGTTC	9119
Db	9061	TGCTTCAGGAACTTGGGTACACCTTGGAACTGGAGACATCGGGCCAGAGTGTTC	9120
Qy	9120	CGCGCTAAGCTACTCTCCAGGGGGAGGGCGGACATTTGGGCAGATACCTCTTTAC	9179
Db	9121	CGCGCTAAGCTACTCTCCAGGGGGAGGGCGGACATTTGGGCAGATACCTCTTTAC	9180
Qy	9180	TGGGCAGTTAAGACCAAGCTTAACTCACTCCAAATCCCGGCGCGTCCAGCTGGACTTG	9239
Db	9181	TGGGCAGTTAAGACCAAGCTTAACTCACTCCAAATCCCGGCGCGTCCAGCTGGACTTG	9240
Qy	9240	TCCTGGCTGTTCTCGTCCCTGGTTACAGGGGGAGACATATATCAGAGCTGTCTGTGCC	9299
Db	9241	TCCTGGCTGTTCTCGTCCCTGGTTACAGGGGGAGACATATATCAGAGCTGTCTGTGCC	9300
Qy	9300	CGACCCCGCTGGTTTCGGTTGTCCTACTCTCTACTTCTGTAGGGTAGGCATTTAAGCTG	9359
Db	9301	CGACCCCGCTGGTTTCGGTTGTCCTACTCTCTACTTCTGTAGGGTAGGCATTTAAGCTG	9360
Qy	9360	CTCCCAACCGATGAACGGGAGGTAACCACTCCAGGCCCTTAAGCCATTTCTGTTTTTT	9419
Db	9361	CTCCCAACCGATGAACGGGAGGTAACCACTCCAGGCCCTTAAGCCATTTCTGTTTTTT	9420

[illegible]

[illegible]

1634 CTCCCTCBAACCTGGGTTCTTTCCGCGCTGTTTACGACACAAGATTCAAATCGTCCGG 1693
 1624 CTCCCTCCAACTGGGTTCTTTCCGCGCTGTTTACGACACAAGATTCAAATCGTCCGG 1683
 1694 GTCCCGGAGCGATGCGCAGCTGCGCCGCCCAATGACTGTTCCGCCAGGGGTGGGGCCC 1753
 1684 GTGCCGGAGCGATGGCAGCTGCGCCGCCCAATGACTGTTCCGCCAGGGGTGGGGCCC 1743
 1754 CATCACTATACATPAGCTAACAGCTCGGATCAGAGCCCTTATGTGGCAATTAACGGCC 1813
 1744 CATCACTATACATPAGCTAACAGCTCGGATCAGAGCCCTTATGTGGCAATTAACGGCC 1803
 1814 TCACACCTGTGCTGCTACACCGCTGCGAGGTGTGTGCTCCAGTGTATTTGTTTACCCC 1873
 1804 TCACACCTGTGCTGCTACACCGCTGCGAGGTGTGTGCTCCAGTGTATTTGTTTACCCC 1863
 1874 AAGCCCTGTGTGTGGGGACCAACCGATCGTTCCGTTGCTCCATAGTATAGCTGGGGGA 1933
 1864 AAGCCCTGTGTGTGGGGACCAACCGATCGTTCCGTTGCTCCATAGTATAGCTGGGGGA 1923
 1934 GAATGACACAGACGTGATGCTCTCAACACACACGCTCCGCCCAAGGCAACTGTTTCGG 1993
 1924 GAATGACACAGACGTGATGCTCTCAACACACACGCTCCGCCCAAGGCAACTGTTTCGG 1983
 1994 CTGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2053
 1984 CTGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2043
 2054 CGGGGGGGTCCGGTAACCGCACTTGAATGTGCCCAACGAGCTGCTTCGGGAAGCAACCCGA 2113
 2044 CGGGGGGGTCCGGTAACCGCACTTGAATGTGCCCAACGAGCTGCTTCGGGAAGCAACCCGA 2103
 2114 GGCCTACTTACACAAATGTGCTCGGGGCTGCTGACACTAGTGCCTAGTAGACTA 2173
 2104 GGCCTACTTACACAAATGTGCTCGGGGCTGCTGACACTAGTGCCTAGTAGACTA 2163
 2174 CCCATACAGGCTTTGGCACTACCCCTGCACCTCAATTTTCCATCTTTAAGTTAGAT 2233
 2164 CCCATACAGGCTTTGGCACTACCCCTGCACCTCAATTTTCCATCTTTAAGTTAGAT 2223
 2234 GTATGTGGGGGGTGGAGCAGAGCTCAATGCCCATGCAATTCGATCGAGGAGAGCG 2293
 2224 GTATGTGGGGGGTGGAGCAGAGCTCAATGCCCATGCAATTCGATCGAGGAGAGCG 2283
 2294 CTGTAACTTGGAGACAGGATAGCTCAGACTCAGCCGCTGCTGTCTACACACAGA 2353
 2284 CTGTAACTTGGAGACAGGATAGCTCAGACTCAGCCGCTGCTGTCTACACACAGA 2343
 2354 GTGGCAGATGCTGCTGCTGCTTACACACCTTACCGGCTTTATCCACTGTTTATCCA 2413
 2344 GTGGCAGATGCTGCTGCTGCTTACACACCTTACCGGCTTTATCCACTGTTTATCCA 2403
 2414 TCTCCATCAGACATCGTGGAGTGCAATACCTGTAGCTGAGGTAGGTGAGCTTGTCTC 2473
 2404 TCTCCATCAGACATCGTGGAGTGCAATACCTGTAGCTGAGGTAGGTGAGCTTGTCTC 2463
 2474 CTTTGCAATCAATGGAGTGATGCTGCTGCTGCTTCTCCCTGGCAGACGCGCGCT 2533
 2464 CTTTGCAATCAATGGAGTGATGCTGCTGCTGCTTCTCCCTGGCAGACGCGCGCT 2523
 2534 GTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2593
 2524 GTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2583
 2594 GGTGGTCTCAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2653
 2584 GGTGGTCTCAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2643
 2654 CTTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2713
 2644 CTTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2703

2714 TGGCGTATGGCGCTGCTCCTCTACTGCGCTTACCAACAGAGCTTACGCTTGA 2773
 2704 TGGCGTATGGCGCTGCTCCTCTACTGCGCTTACCAACAGAGCTTACGCTTGA 2763
 2774 CCGGAGATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2833
 2764 CCGGAGATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2823
 2834 GTCAACATATCAAAAGTGTCTCTACTAGGCTCATATGCTGCTTACATATCTTATCAC 2893
 2824 GTCAACATATCAAAAGTGTCTCTACTAGGCTCATATGCTGCTTACATATCTTATCAC 2883
 2894 CAGAGCGAGGCGCATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2953
 2884 CAGAGCGAGGCGCATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2943
 2954 TGCCATCATCTCTCTCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3013
 2944 TGCCATCATCTCTCTCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3003
 3014 CTTGCTGCGCATACTCGGCCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3073
 3004 CTTGCTGCGCATACTCGGCCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3063
 3074 CTTGCTGCGCGCTCAAGGGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3133
 3064 CTTGCTGCGCGCTCAAGGGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3123
 3134 TCATTATGTCCTCAATGGTCTTCAAGCTGCGGCGCTGACAGGTACGTACGTTATTA 3193
 3124 TCATTATGTCCTCAATGGTCTTCAAGCTGCGGCGCTGACAGGTACGTACGTTATTA 3183
 3194 CCATCTTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3253
 3184 CCATCTTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3243
 3254 AGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3313
 3244 AGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3303
 3314 TGCTGTGGGACATCATCTTGGCTTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3373
 3304 TGCTGTGGGACATCATCTTGGCTTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3363
 3374 TTTGGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3433
 3364 TTTGGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3423
 3434 CTACTCCCAACAAACGCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3493
 3424 CTACTCCCAACAAACGCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3483
 3494 CAAGAACAGGCTGCAAGGGGAGGTTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3553
 3484 CAAGAACAGGCTGCAAGGGGAGGTTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3543
 3554 GACCTGCTCAACAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3613
 3544 GACCTGCTCAACAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3603
 3614 CGGTCCCAAAAGGCTCAAAATGACCAAAATGACCAAAATGACCAAAATGACCAAAATGACCAAAATGACCAAAAT 3673
 3604 CGGTCCCAAAAGGCTCAAAATGACCAAAATGACCAAAATGACCAAAATGACCAAAATGACCAAAATGACCAAAAT 3663
 3674 GCAGGCGCCCCCGGGGCGGCTCCATGACACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3733
 3664 GCAGGCGCCCCCGGGGCGGCTCCATGACACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3723
 3734 CTTTGTGACAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3793
 3724 CTTTGTGACAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3783
 3794 TCTACTCTCCCCCAGGCGCGCTCTCTCTTACCTGAAAGGCTCTCTCGGGTGGTCCATTTGCTTTG 3853

QY	6014	CTTGGTCTGGGGTCTGTGCGGCAAAATACCTGCTCGGACGTGGGCGCGGAGAGGG	6073	QY	7094	AATTTCTGGAATCTTTTGGAAACCGCTTCAACGCGAGGGGATGAGAGGAGATATCCCTGCG	7153
Db	6004	CTTGGTCTGGGGTCTGTGCGGCAAAATACCTGCTCGGACGTGGGCGCGGAGAGGG	6063	Db	7084	AATTTCTGGAATCTTTTGGAAACCGCTTCAACGCGAGGAGATATCCCTGCG	7143
QY	6074	GGCTGTGAGTGGATGAACCGCTGTAGAGGTTCGCTTCGGGGGTAAACCACTCTCCCC	6133	QY	7154	GGCGGAGATCTTGGGAAATCCAGGAATTTCCCTCAGCGTTGCTCCCATATGGGCAAGCC	7213
Db	6064	GGCTGTGAGTGGATGAACCGCTGTAGAGGTTCGCTTCGGGGGTAAACCACTCTCCCC	6123	Db	7144	GGCGGAGATCTTGGGAAATCCAGGAATTTCCCTCAGCGTTGCTCCCATATGGGCAAGCC	7203
QY	6134	TACGCACTATGTGCTGAGAGCGACGCTGAGAGGTTCGCTTCGGGGGTAAACCACTCTCCCC	6193	QY	7214	GGACTAATATCTTCCCTAGAGTCTTGGGAGGACCGGAGTACGTTCCCTCCGCTGGT	7273
Db	6124	CACGCACTATGTGCTGAGAGCGACGCTGAGAGGTTCGCTTCGGGGGTAAACCACTCTCCCC	6183	Db	7204	GGACTAATATCTTCCCTAGAGTCTTGGGAGGACCGGAGTACGTTCCCTCCGCTGGT	7263
QY	6194	TACCACTACTCACTGCTGAGAGCGCTCCACAGTGGATTAAATGAGGAGTCTCTACGCC	6253	QY	7274	ACGCGGATCCCATTTGCACTACCAAGCTCTCCCAATACCACTCCACGAGAAAGAG	7333
Db	6184	TACCACTACTCACTGCTGAGAGCGCTCCACAGTGGATTAAATGAGGAGTCTCTACGCC	6243	Db	7264	ACGCGGATCCCATTTGCACTACCAAGCTCTCCCAATACCACTCCACGAGAAAGAG	7323
QY	6254	ATGCTCCGGCTCGTGGCTAAGGATTTTGGGATTTGGATATGACAGGTGTTGACTGACTT	6313	QY	7334	GACGGTGTCTCTGACAGAAATCCAAATGTCTTCTGCTTGGCGAGCTCCCACTAAGAC	7393
Db	6244	ATGCTCCGGCTCGTGGCTAAGGATTTTGGGATTTGGATATGACAGGTGTTGACTGACTT	6303	Db	7324	GACGGTGTCTCTGACAGAAATCCAAATGTCTTCTGCTTGGCGAGCTCCCACTAAGAC	7383
QY	6314	CAAGACTCTGGCTCCAGTCCAAACTCTGCGCGGTTACCGGAGTCCCTTTCTCTGATG	6373	QY	7394	CTTCCGTAGCTCCGANTCTGCGCGTGTGATAGCGGACCGGACCGCTTCTCTGACCT	7453
Db	6304	CAAGACTCTGGCTCCAGTCCAAACTCTGCGCGGTTACCGGAGTCCCTTTCTCTGATG	6363	Db	7384	CTTCCGTAGCTCCGANTCTGCGCGTGTGATAGCGGACCGGACCGCTTCTCTGACCT	7443
QY	6374	CCAAACCGGCTACAGGAGTCTGCGGGGGGACCGGATCATGCAAAACCACTGCCCCATG	6433	QY	7454	GGCTCCGACGACGCTGCAAAAGATCCGAGTTGAGTCTGCTACTCTCTCCATGCCCTCT	7513
Db	6364	CCAAACCGGCTACAGGAGTCTGCGGGGGGACCGGATCATGCAAAACCACTGCCCCATG	6423	Db	7444	GGCTCCGACGACGCTGCAAAAGATCCGAGTTGAGTCTGCTACTCTCTCCATGCCCTCT	7503
QY	6434	CGGAGACAGATCGCGGACATGTCAAAACCGTTTCCATGAGATCGTAGGCGCTAGAAC	6493	QY	7514	TEAAGGGGAGCGGGGGACCCCGATCTCAGCGAGGGCTTTGCTGTCTACCTGAGTGAAGA	7573
Db	6424	CGGAGACAGATCGCGGACATGTCAAAACCGTTTCCATGAGATCGTAGGCGCTAGAAC	6483	Db	7504	TEAAGGGGAGCGGGGGACCCCGATCTCAGCGAGGGCTTTGCTGTCTACCTGAGTGAAGA	7563
QY	6494	CTGAGACACAGTGGGACGGAAGCTTCCCATCAACGATACACACCGGACCTTGAC	6553	QY	7574	GGCTAGTGAGATGTCTGCT	7633
Db	6484	CTGAGACACAGTGGGACGGAAGCTTCCCATCAACGATACACACCGGACCTTGAC	6543	Db	7564	GGCTAGTGAGATGTCTGCT	7623
QY	6554	ACCTTCCCGGCGCCAACTATTCAGGCGCTATGCGGGGTGGCTGTGAGGATACGT	6613	QY	7634	GCCATGCGCTCGGAGGAAAGTAAAGCTCCCATCAACCGCTTACCACTCTCTCTCTCT	7693
Db	6544	ACCTTCCCGGCGCCAACTATTCAGGCGCTATGCGGGGTGGCTGTGAGGATACGT	6603	Db	7624	GCCATGCGCTCGGAGGAAAGTAAAGCTCCCATCAACCGCTTACCACTCTCTCTCTCT	7683
QY	6614	GGAGTTACCGTGTGGGGATTTCCACTACGTCAGCGGATGACCACTGACAACTATAA	6673	QY	7694	TCACCAACATCTGTCTACGCGACCAACATCCCGAGCGGACCTCCCGGAGAGAGGT	7753
Db	6604	GGAGTTACCGTGTGGGGATTTCCACTACGTCAGCGGATGACCACTGACAACTATAA	6663	Db	7684	TCACCAACATCTGTCTACGCGACCAACATCCCGAGCGGACCTCCCGGAGAGAGGT	7743
QY	6674	GTCCCATGCGAGTTCCCGGCGCCCGGATTTTCCAGGAGTGGATGGAGTGGCTGCA	6733	QY	7754	CACCTTTGACATTTGGAAGTCTGAGTATCATTTACCGGAGCTACTCAAGAGATGAA	7813
Db	6664	GTCCCATGCGAGTTCCCGGCGCCCGGATTTTCCAGGAGTGGATGGAGTGGCTGCA	6723	Db	7744	CACCTTTGACATTTGGAAGTCTGAGTATCATTTACCGGAGCTACTCAAGAGATGAA	7803
QY	6734	CAGGTACGCTCCGGCGTGCAAAACCTCTTACGCGGAGAGCTCACGTTCCAGGTCCGGCT	6793	QY	7814	GGCGAGGCGTCCACAGTTTAAAGCTTCTATCTATAGAGGAGCTGCAAGTGCAC	7873
Db	6724	CAGGTACGCTCCGGCGTGCAAAACCTCTTACGCGGAGAGCTCACGTTCCAGGTCCGGCT	6783	Db	7804	GGCGAGGCGTCCACAGTTTAAAGCTTCTATCTATAGAGGAGCTGCAAGTGCAC	7863
QY	6794	CAACCAATCTTGGTCCGGTCCGAGTCCCATGCGGCGCCGACCGGACGTAACAGTCT	6853	QY	7874	GCCCCACATTTCCGCAAAATTTGGCTATGCGGCAAGAGAGCTCCGGAACCTATC	7933
Db	6784	CAACCAATCTTGGTCCGGTCCGAGTCCCATGCGGCGCCGACCGGACGTAACAGTCT	6843	Db	7864	GCCCCACATTTCCGCAAAATTTGGCTATGCGGCAAGAGAGCTCCGGAACCTATC	7923
QY	6854	TACTTCAATGCTTACCGATCCCTCCGATTTACAGAGAGAGCGCTAAGCTGAGCTGGC	6913	QY	7934	CAGAGGCGGCTTAAACCACTCCGCTCCGCTGCGGAGAGCTTCTCTGGAAGACACTGAAAC	7993
Db	6844	TACTTCAATGCTTACCGATCCCTCCGATTTACAGAGAGAGCGCTAAGCTGAGCTGGC	6903	Db	7924	CAGAGGCGGCTTAAACCACTCCGCTCCGCTGCGGAGAGCTTCTCTGGAAGACACTGAAAC	7983
QY	6914	TAGAGGGTCTCCCGCTTATAGCGAGCTCATCAGCTAGCGAGTGTGCTGCGCTCTCTT	6973	QY	7994	ACCAATTTGACACCACTCATGCGCAAAAGTGGAGTCTTCTGCTCCCAACAGAGAGGG	8053
Db	6904	TAGAGGGTCTCCCGCTTATAGCGAGCTCATCAGCTAGCGAGTGTGCTGCGCTCTCTT	6963	Db	7984	ACCAATTTGACACCACTCATGCGCAAAAGTGGAGTCTTCTGCTCCCAACAGAGAGGG	8043
QY	6974	GAAGGCGACATGCACTACCCACCACTACTCCCGGAGCGTCACTCACTGAGGCGCAACT	7033	QY	8054	AGGCGGCAAGCGAGCTCGCTTATCGTATTTCCAGAGCTGGGAGTCTGCTATGCGGAGAA	8113
Db	6964	GAAGGCGACATGCACTACCCACCACTACTCCCGGAGCGTCACTCACTGAGGCGCAACT	7023	Db	8044	AGGCGGCAAGCGAGCTCGCTTATCGTATTTCCAGAGCTGGGAGTCTGCTATGCGGAGAA	8103
QY	7034	CTTGTGGGCGGAGGAGTGGGCGGAAACATCACTCCGCTGGAGTCAAGAGATAGGTAGT	7093	QY	8114	GATGCGCTTTTACGACGCTGCTCCACCTTCTCCAGCGCTGATGGCTCTCTCATACGG	8173
Db	7024	CTTGTGGGCGGAGGAGTGGGCGGAAACATCACTCCGCTGGAGTCAAGAGATAGGTAGT	7083	Db	8104	GATGCGCTTTTACGACGCTGCTCCACCTTCTCCAGCGCTGATGGCTCTCTCATACGG	8163
				QY	8174	ATTTCAATACTCTCCCCAAGAGCGGGTTCGAGTTCTCTGCTGAATACCTGGAAATCAAGAA	8233

Db	8164	RTTTCATACTCCCGCAAGCGGGTTCAGTTCCTGCTGATACCTGGAATCAAGAA	8223	Db	9244	CGCTGTTACAGCGGGGAGACATATATCAAGCCTGTCTCGTCCGACCCCGCTGTT	9303
Qy	8234	ATGCCCTATGCGGTTCTCATATGACACCCGCTGTTTGTACTCAACGGTCACTGAGAGTGA	8293	Qy	9314	TCCGTTGTCCTACTCTTACTTCTGTAGGGGTAGGCAATTTACCTGCTCCCCACCGATG	9373
Db	8224	ATGCCCTATGCGGTTCTCATATGACACCCGCTGTTTGTACTCAACGGTCACTGAGAGTGA	8283	Db	9304	TCCGTTGTCCTACTCTTACTTCTGTAGGGGTAGGCAATTTACCTGCTCCCCACCGATG	9363
Qy	8294	CATTCGTGTGAGGAGTCAATTTACCAATGTTGTGACTTGGCCCGCCGAGGCCACAGAGC	8353	Qy	9374	AACGGGGAGCTAACCACTCCAGGCTTAAGCCATTTCCTGTTTTTTTTTTTTTTTTTTT	9433
Db	8284	CATTCGTGTGAGGAGTCAATTTACCAATGTTGTGACTTGGCCCGCCGAGGCCACAGAGC	8343	Db	9364	AACGGGGAGCTAACCACTCCAGGCTTAAGCCATTTCCTGTTTTTTTTTTTTTTTTTTT	9423
Qy	8354	CATAAGGTCCCTCACAGAGCGGCTTACATCGCGGGTCCCTGACTAACTCAAAAAGGGA	8413	Qy	9434	TTT	9493
Db	8344	CATAAGGTCCCTCACAGAGCGGCTTACATCGCGGGTCCCTGACTAACTCAAAAAGGGA	8403	Db	9424	TTT	9483
Qy	8414	GAATCGGGTTATTCGCGGTCGCGCGCAAGTGGCGTCTGACGACTAGCTGCGGTAAATAC	8473	Qy	9494	T 9494	
Db	8404	GAATCGGGTTATTCGCGGTCGCGCGCAAGTGGCGTCTGACGACTAGCTGCGGTAAATAC	8463	Db	9484	T 9484	
Qy	8474	CCTCAGATGTTACTTGAAGGCCACTGTCAGCCTGTCGAGCTGCAAGCTCCAGGACTGCAC	8533				
Db	8464	CCTCAGATGTTACTTGAAGGCCACTGCGGCTGTCGAGCTGCAAGCTCCAGGACTGCAC	8523				
Qy	8534	GATCCTCGTCAACGAGACAGCTTGTGCTTATCTGTGAAGCGCGGAAACCCAGGAGGA	8593				
Db	8524	GATCCTCGTCAACGAGACAGCTTGTGCTTATCTGTGAAGCGCGGAAACCCAGGAGGA	8583				
Qy	8594	TGCGGCGCCCTACGAGCCTTCAACGAGGCTATGACTAGGTATTCGCCCGCCCGGGGA	8653				
Db	8584	TGCGGCGCCCTACGAGCCTTCAACGAGGCTATGACTAGGTATTCGCCCGCCCGGGGA	8643				
Qy	8654	TCCGCCCAACGAAATACGACCTGGAGCTGATAACATCATGTTTCCTCAATGTGTGAGT	8713				
Db	8644	TCCGCCCAACGAAATACGACCTGGAGCTGATAACATCATGTTTCCTCAATGTGTGAGT	8703				
Qy	8714	CGCGCACGATCATCTGCAAAAGGGTATCTACTCACCGTGAACCCCAACCCCGCT	8773				
Db	8704	CGRCACGATCATCTGCAAAAGGGTATCTACTCACCGTGAACCCCAACCCCGCT	8763				
Qy	8774	TGCAGGGCTGCTGGGAGACAGCTAGACACACTCCATCACTTGGCTAGGCAAT	8833				
Db	8764	TGCAGGGCTGCTGGGAGACAGCTAGACACACTCCATCACTTGGCTAGGCAAT	8823				
Qy	8834	CATCATGTATGCGCCACCTATGCGGAGGATGATTTGTGATGACTCACTTTTCTCCAT	8893				
Db	8824	CATCATGTATGCGCCACCTATGCGGAGGATGATTTGTGATGACTCACTTTTCTCCAT	8883				
Qy	8894	CCTTCTAGCTCAAGAGCAACTTGAAGAGCCCTGATGTCAGATCTACGGGCTTGCTA	8953				
Db	8884	CCTTCTAGCTCAAGAGCAACTTGAAGAGCCCTGATGTCAGATCTACGGGCTTGCTA	8943				
Qy	8954	CTCCATTGAGCACTTGACCTACCTAGATCATGACGACCCCATGCTTGTAGCGCAT	9013				
Db	8944	CTCCATTGAGCACTTGACCTACCTAGATCATGACGACCCCATGCTTGTAGCGCAT	9003				
Qy	9014	TACACTCCACAGTACTCTCCAGGTGAGATCAATAGGGTGGCTTCATGCTCAGGAAT	9073				
Db	9004	TACACTCCACAGTACTCTCCAGGTGAGATCAATAGGGTGGCTTCATGCTCAGGAAT	9063				
Qy	9074	TGGGGTACCACCTTCGAACTTGGAGACATCGGGCCAGAGTGTCCGCGCTAAGTACT	9133				
Db	9064	TGGGGTACCACCTTCGAACTTGGAGACATCGGGCCAGAGTGTCCGCGCTAAGTACT	9123				
Qy	9134	GTCCAGGGGGAGGGCCGCACTTGTGGCAGATACCTTTTAACTGGGACGTAAGGAC	9193				
Db	9124	GTCCAGGGGGAGGGCCGCACTTGTGGCAGATACCTTTTAACTGGGACGTAAGGAC	9183				
Qy	9194	CAGAGCTTAACTCACTCCATCCCGCGGCTCCAGCTTGAAGTGTGTGCTGGCTTCGT	9253				
Db	9184	CAGAGCTTAACTCACTCCATCCCGCGGCTCCAGCTTGAAGTGTGTGCTGGCTTCGT	9243				
Qy	9254	CGTGGTTACAGCGGGGGAGACATATATACAGCCTGTCTGCTGCCGACCCCGCTGTT	9313				

LOCUS	HPGJ491
DEFINITION	Hepatitis C virus genome, complete sequence.
ACCESSION	D10750.1
VERSION	GI:221606
KEYWORDS	C protein; E protein; E2 protein; NS1 protein; NS2 protein; NS3 protein; NS4 protein; NS5 protein.
SOURCE	Hepatitis C virus
ORGANISM	Hepatitis C virus
REFERENCE	1 (bases 1 to 9460) Okamoto,H., Kojima,M., Okada,S., Yoshizawa,H., Iizuka,H., Tanaka,T., Machimori,E.E., Peterson,D.A., Ito,Y. and Mishiro,S. Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability Virology 190 (2), 894-899 (1992)
TITLE	
JOURNAL	
MEDLINE	92391112
PUBMED	1325713
REFERENCE	2 (bases 1 to 9460) Okamoto,H. Direct Submission Submitted (18-MAR-1992) Hiroaki Okamoto, Jichi Medical School, Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi 329-04, Japan (E-mail:hokamoto@jichi.ac.jp) Tel:0285-44-2111(ex.334), Fax:0285-44-1557) Submitted (18-Mar-1992) to DDBJ by:
TITLE	
JOURNAL	
COMMENT	Hiroaki Okamoto Immunology Division Jichi Medical School Kawachi-gun, Tochigi 329-04 Japan Phone: 0285-44-2111 x3334 Fax: 0285-44-1557.
FEATURES	Location/Qualifiers
source	1. .9460 /organism="Hepatitis C virus" /mol_type="genomic RNA" /strain="HC-J4" /isolate="HC-J4/91" /db_xref="taxon:11103" 342..9374 /codon_start=1 /product="polyprotein precursor" /protein_id="BAA01383.1" /db_xref="GI:221607"
CDS	/translation="MSTNFKPQRKTKNTNRPDQVKFGGGQIVGGVYLLPRRGRL GVATKTSRSPQRRGRQPIPKARQPEGRAWAQGYWPFLYNGSELGWAGMLSPRG SRPSWPTDRSRNLGVIDLTCTGFPADLMGYTFLVGAPLGGARALAHGVRVLED GVNYAGNLPCCSFSLFLLALISCLTIPASAYEVRNVSIGIYHVTNDCNSSIYFAEA VIYHTPCVPLVREGSSRCWALPTFLAARDASVPTTIRHVDLLVCTAAFCGAMY VGDLCSIFLVSQLTFSRRHETVQDCNSIYPGVSHRWMDMMKWSPTTLLV SQLRLTPQAVVDVAGAHWGLAGLAYISWGNWAKVLIVALLFAGVDGATYTSGGVA

RESULT 6
 HPCJ491
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 1 (bases 1 to 9460)
 Okamoto,H., Kojima,M., Okada,S., Yoshizawa,H., Iizuka,H., Tanaka,T., Muchmore,E.E., Peterson,D.A., Ito,Y. and Mishiro,S.
 Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability
 Virology 190 (2), 894-899 (1992)
 92391112
 1325713
 REFERENCE
 2 (bases 1 to 9460)
 Okamoto,H.
 Direct Submission
 Submitted (18-Mar-1992) Hiroaki Okamoto, Jichi Medical School, Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi 329-04, Japan (E-mail:hokamoto@jichi.ac.jp, Tel:0285-44-2111(ex:3334), Fax:0285-44-1557)
 Submitted (18-Mar-1992) to DDBJ by:
 Hiroaki Okamoto
 Immunology Division
 Jichi Medical School
 Kawachi-gun, Tochigi 329-04
 Japan
 Phone: 0285-44-2111 x3334
 Fax: 0285-44-1557
 Location/Qualifiers
 1.9460
 /organism="Hepatitis C virus"
 /mol_type="genomic RNA"
 /strain="HC-J4"
 /isolate="HC-J4/91"
 /db_xref="taxon:11103"
 342_9374
 /codon_start=1
 /product="polyprotein precursor"
 /protein_id="BAA01583.1"
 /translation="MSTNPKPQRTKRNTRRPQDKPFGGQIVGGVYLLPRRGRPLGVATRKTSERSQPRGRQPIKAPQPSGRAWQPGWPPLYNGLGAWGLLSPRGSRPSGPTDPRRSRLNKGVIDTLTQGPADLMGTPLVGAPLGGAAALAGVRLEDGVNATGNDLPFGCSIFLLALSCLTIPASAYEVNVSIGYHVTDNCSNLSIVYAADVTMHTPGCVPCVRGNRSRCAVUTFLAARDASVPTTIRHVDLLVGTAFRCASYVGDLCGSIFVQLFTFSRPHRTYQDCNSCIYPGHVSGRHNMMDMMNWSPTTALVVSQLLRI PQAVVDNVAGHWGLAGLAVYSVMGNKVLIVALLFAGVDGATYTSGEVA

FEATURES
 source
 CDS

1141 GGACGGCTGCTTTCTGCTCCGCTATGATACGAGGAGATCTCTCGGATCTATTTCTCCG 1200
 1141 GGACGGCTGCTTTCTGCTCCGCTATGATACGAGGAGATCTCTCGGATCTATTTCTCCG 1200
 1201 TCTCCAGCTGTTCACCTTCTGCTCCGCTCCGCGCATGAGACAGTGCAGGACTGCAACTGCT 1260
 1201 TCTCCAGCTGTTCACCTTCTGCTCCGCTCCGCGCATGAGACAGTGCAGGACTGCAACTGCT 1260
 1261 CAATCTATCCCGGCAATGATCAGGTACCGCATGGCTTGGGATGATGATGAACCTGGT 1320
 1261 CAATCTATCCCGGCAATGATCAGGTACCGCATGGCTTGGGATGATGATGAACCTGGT 1320
 1321 CACCTACACAGCCCTAGTGTGTGCGCATGCTCCGATCCCAAGCTGTGCGGACA 1380
 1321 CACCTACACAGCCCTAGTGTGTGCGCATGCTCCGATCCCAAGCTGTGCGGACA 1380
 1381 TGGTGGCGGGGCGGACCTGCGGAGTCTTGGCGGCTTGCCTACTATTCCATGGTAGGGA 1440
 1381 TGGTGGCGGGGCGGACCTGCGGAGTCTTGGCGGCTTGCCTACTATTCCATGGTAGGGA 1440
 1441 ACTGGGCTAAGTGTCTGATTTGCGGCTACTCTTTGCCGGGTTGACGGGGACCCACA 1500
 1441 ACTGGGCTAAGTGTCTGATTTGCGGCTACTCTTTGCCGGGTTGACGGGGACCCACA 1500
 1501 CGACGGGGAGGTGGCGGGGCAACACCTCCGGGTTCACTCTCTTTCTCATCTGGGG 1560
 1501 CGTCCGGGGGGGTGGCGGGGCGCACACCTCCGGGTTCACTCTCTTTCTCATCTGGGG 1560
 1561 GGTCTCAGAAAATCCAGCTTGTGAAATACCAAGCGAGCTGSCACATCAACAGAGCTGCC 1620
 1561 GGTCTCAGAAAATCCAGCTTGTGAAATACCAAGCGAGCTGSCACATCAACAGAGCTGCC 1620
 1621 TAAATTCGAATGACTCCCTCAAACCTGGGTTCTTTGCGGCTGTTTACGACACAAAGT 1680
 1621 TAAATTCGAATGACTCCCTCAAACCTGGGTTCTTTGCGGCTGTTTACGACACAAAGT 1680
 1681 TCAACTGCTCGGGTGGCGGGGCGGATGGCGAGCTGCGGCGGCTGATGCTGCTGCTCC 1740
 1681 TCAACTGCTCGGGTGGCGGGGCGGATGGCGAGCTGCGGCGGCTGATGCTGCTGCTCC 1740
 1741 AGGGTGGGGGCGGATCACTATTAAGCTTAACAGCTGAGTCAAGGCTGAGGCTTATGCT 1800
 1741 AGGGTGGGGGCGGATCACTATTAAGCTTAACAGCTGAGTCAAGGCTGAGGCTTATGCT 1800
 1801 GGCATGAGCGGCTGAGCGGCTGAGTCAAGGCTTACAGCGGCTGAGGCTGAGGCTTAC 1860
 1801 GGCATGAGCGGCTGAGCGGCTGAGTCAAGGCTTACAGCGGCTGAGGCTGAGGCTTAC 1860
 1861 ATTGTTTCAACCCAGCGCTGTTGTTGGGGGACACCGATGCTTCCGGTGTCCCTAGT 1920
 1861 ATTGTTTCAACCCAGCGCTGTTGTTGGGGGACACCGATGCTTCCGGTGTCCCTAGT 1920
 1921 ATAGCTGGGGGAGGATGAGACAGAGCTGATGCTCTCAACAAACAGCGGCTGCGGACAAG 1980
 1921 ATAGCTGGGGGAGGATGAGACAGAGCTGATGCTCTCAACAAACAGCGGCTGCGGACAAG 1980
 1981 GCAATGTTGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
 1981 GCAATGTTGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
 2041 CCCCGTGAACATCGGGGGGCTGCTTAACCGGACCTTGTATCTGCTCCCGGAGCTGCTTCC 2100
 2041 CCCCGTGAACATCGGGGGGCTGCTTAACCGGACCTTGTATCTGCTCCCGGAGCTGCTTCC 2100
 2101 GGAAGCACCCCGAGGCTACTTACAAAATGTTGGCTCGGGGCGCTGTTGACACCTAGT 2160
 2101 GGAAGCACCCCGAGGCTACTTACAAAATGTTGGCTCGGGGCGCTGTTGACACCTAGT 2160
 2161 GCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 2161 GCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 2221 TTAAAGTTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280

2221 TTAAAGTTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 2281 CTGAGGAGAGCCCTGTAACTTTGAGGACAGGATAGGTGAGAACTAGCCCTGCTGTC 2340
 2281 CTGAGGAGAGCCCTGTAACTTTGAGGACAGGATAGGTGAGAACTAGCCCTGCTGTC 2340
 2341 TGTCTACACAGAGAGTGGCAGATAGTCCCTGTGCTTTCACCACTTACCGGCTTTATCCA 2400
 2341 TGTCTACACAGAGAGTGGCAGATAGTCCCTGTGCTTTCACCACTTACCGGCTTTATCCA 2400
 2401 CTGTTTGTATTCATCTCCATCAGAACATCGTGACGTGCAATACCTTACCGGTGAGGT 2460
 2401 CTGTTTGTATTCATCTCCATCAGAACATCGTGACGTGCAATACCTTACCGGTGAGGT 2460
 2461 CAGCGTGTGCTCTCTTTCGAAATCAAAATGGGAGTACATCTCTGTTCTTCTCTCTCTGG 2520
 2461 CAGCGTGTGCTCTCTTTCGAAATCAAAATGGGAGTACATCTCTGTTCTTCTCTCTCTGG 2520
 2521 CAGACGCGCGGTGTGCTCTGTTGTTGATGATGCTGCTGATAGCCAGGCTCAGGCGG 2580
 2521 CAGACGCGCGGTGTGCTCTGTTGTTGATGATGCTGCTGATAGCCAGGCTCAGGCGG 2580
 2581 CTTTACAGAACTTGTGTTGCTCTCAATGCGGCTCGGTGCGGAGCGCATGTTCTCT 2640
 2581 CTTTACAGAACTTGTGTTGCTCTCAATGCGGCTCGGTGCGGAGCGCATGTTCTCT 2640
 2641 CTTTCTTGTGTTCTTCTGCGGCTCTGTATTAAGGAGGAGGCTGCTCTGCGGCGG 2700
 2641 CTTTCTTGTGTTCTTCTGCGGCTCTGTATTAAGGAGGAGGCTGCTCTGCGGCGG 2700
 2701 CGTATGCTTTTATGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
 2701 CGTATGCTTTTATGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
 2761 CTTACGCTTTGACCGGAGATGGTGCATCGTGGGGGGTGGGTTCTGTAGTCTG 2820
 2761 CTTACGCTTTGACCGGAGATGGTGCATCGTGGGGGGTGGGTTCTGTAGTCTG 2820
 2821 TATTTCTGAGCTTGTACCATACATAAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
 2821 TATTTCTGAGCTTGTACCATACATAAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
 2881 AATATCTTATCACAGAGCGGAGCGGACAGTGTGGGTGCGGCTCCCGCTCAACGTTT 2940
 2881 AATATCTTATCACAGAGCGGAGCGGACAGTGTGGGTGCGGCTCCCGCTCAACGTTT 2940
 2941 GGGGAGCGCGGATGCCATCT 3000
 2941 GGGGAGCGCGGATGCCATCT 3000
 3001 ACATFCAACAACT 3060
 3001 ACATFCAACAACT 3060
 3061 CAGAGTGGCGTCT 3120
 3061 CAGAGTGGCGTCT 3120
 3121 AAGTGGCGCGGGGTCAATATGTCAAAATGTTTCTATGAAGCTGGGCGGCTGACAGTA 3180
 3121 AAGTGGCGCGGGGTCAATATGTCAAAATGTTTCTATGAAGCTGGGCGGCTGACAGTA 3180
 3181 CTTAGCTTTTAAACATCTTACCCCATCTGGGAGTGGGCGGCGGCTTACGAGACC 3240
 3181 CTTAGCTTTTAAACATCTTACCCCATCTGGGAGTGGGCGGCGGCTTACGAGACC 3240
 3241 TTGGCGGTGGCGGTAGAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3300
 3241 TTGGCGGTGGCGGTAGAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3300
 3301 GAGCAGACACCGCT 3360

Db 3301 GAGAGACACCGGAGCATGTGGAGACATCATCTTTGGGCTTACCGTCTCTCGGCCGGAAGG 3360
 Qy 3361 GGAAGGAGATATTTTGGGACCGGCTCATAGTCTTCGAAAGGCAAGGTTGGGACTCTCTTG 3420
 Db 3361 GGAAGGAGATATTTTGGGACCGGCTCATAGTCTTCGAAAGGCAAGGTTGGGACTCTCTTG 3420
 Qy 3421 CGCCCATCAGGCTACTCCCAACAAACCGGGGCTACTTGGTTGCATCATCACTAGCC 3480
 Db 3421 CGCCCATCAGGCTACTCCCAACAAACCGGGGCTACTTGGTTGCATCATCACTAGCC 3480
 Qy 3481 TCACAGGCGGCAACAGAACAGGTTCGAAGGGAGGTTCAAGTGGTTTTCACCGCAACAC 3540
 Db 3481 TCACAGGCGGCAACAGAACAGGTTCGAAGGGAGGTTCAAGTGGTTTTCACCGCAACAC 3540
 Qy 3541 AATCTTTCTCGGAGCTGTGCTCAACAGCGCTGTGCTGAGCTGTCTACCATCGCGCTGGCT 3600
 Db 3541 AATCTTTCTCGGAGCTGTGCTCAACAGCGCTGTGCTGAGCTGTCTACCATCGCGCTGGCT 3600
 Qy 3601 CGAAGACCTAGCCGCTCCAAAGGTCCAAATCACCCAAATGTACACCAATGTAGACCTTGG 3660
 Db 3601 CGAAGACCTAGCCGCTCCAAAGGTCCAAATCACCCAAATGTACACCAATGTAGACCTTGG 3660
 Qy 3661 ACCTCGTGGGTGGCAGCGGCCCGCGGCGCTCCATGACACCATGCGAGCTGTGCA 3720
 Db 3661 ACCTCGTGGGTGGCAGCGGCCCGCGGCGCTCCATGACACCATGCGAGCTGTGCA 3720
 Qy 3721 GCTCGGACTTTACTTGGTTCAGAGACATGCTGATTCCTCCGCTGGCGCGCGGAGGCG 3780
 Db 3721 GCTCGGACTTTACTTGGTTCAGAGACATGCTGATTCCTCCGCTGGCGCGCGGAGGCG 3780
 Qy 3781 ACAGCAGGGGAGTCTACTCTCCCGAGCGCGCTCTCTTACCTGAAAGGCTCTCTCGGGTG 3840
 Db 3781 ACAGCAGGGGAGTCTACTCTCCCGAGCGCGCTCTCTTACCTGAAAGGCTCTCTCGGGTG 3840
 Qy 3841 GTCCATGTCTTTGCCCTTCGGGCGAGCTGCTGGGCGCTTCGCGGCTGTGTGTGCAACC 3900
 Db 3841 GTCCATGTCTTTGCCCTTCGGGCGAGCTGCTGGGCGCTTCGCGGCTGTGTGTGCAACC 3900
 Qy 3901 GGGGGTCCGGAAGGGGTGGACTTATACCCGTTGAGTCTATGGAACTACCATCGCGT 3960
 Db 3901 GGGGGTCCGGAAGGGGTGGACTTATACCCGTTGAGTCTATGGAACTACCATCGCGT 3960
 Qy 3961 CTCGGTCTTCACAGAACACTCTATCCCGTTGAGTCTATGGAACTACCATCGCGT 4020
 Db 3961 CTCGGTCTTCACAGAACACTCTATCCCGTTGAGTCTATGGAACTACCATCGCGT 4020
 Qy 4021 ATCTGCACGCTCTACTTGGCAGCGGCAAGAGCACCTAAAGTTCGCGGCTGCGTATGCGAGCC 4080
 Db 4021 ATCTGCACGCTCTACTTGGCAGCGGCAAGAGCACCTAAAGTTCGCGGCTGCGTATGCGAGCC 4080
 Qy 4081 AAGGTTACAGGTGCTGCTCTGAAACCCCTCGCTTGGCGGCTTGGGCTTGGGCGGT 4140
 Db 4081 AAGGTTACAGGTGCTGCTCTGAAACCCCTCGCTTGGCGGCTTGGGCTTGGGCGGT 4140
 Qy 4141 ATATGTCGAAGGCAACAGGTATCGACCTTAACATCAGAACTGGGTTAAGGACCATTAACA 4200
 Db 4141 ATATGTCGAAGGCAACAGGTATCGACCTTAACATCAGAACTGGGTTAAGGACCATTAACA 4200
 Qy 4201 CGGCGGCTCCATTAAGTATCTGACCTTAAGTATCTGAGTATCTGAGTATCTGAGTATCTG 4260
 Db 4201 CGGCGGCTCCATTAAGTATCTGACCTTAAGTATCTGAGTATCTGAGTATCTGAGTATCTG 4260
 Qy 4261 GGGCGGCTATGACATCATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
 Db 4261 GGGCGGCTATGACATCATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
 Qy 4321 TGGGCTATGCGCACAGTCTCTGACCAACAGCGGAGAGCGGTTCGAGCGGCGTCTGCTGCTG 4380
 Db 4321 TGGGCTATGCGCACAGTCTCTGACCAACAGCGGAGAGCGGTTCGAGCGGCGTCTGCTGCTG 4380
 Qy 4381 CCACGCTTACACCTCCGGGATCGGTTACCGTGGCACACCCCAATATGAGGAAATAGGCC 4440
 Db 4381 CCACGCTTACACCTCCGGGATCGGTTACCGTGGCACACCCCAATATGAGGAAATAGGCC 4440

Qy 4441 TGTCACCAATGAGAGATCCCGTCTTATGGCAAGCAATCCCATTTGGGGCATCAAGG 4500
 Db 4441 TGTCACCAATGAGAGATCCCGTCTTATGGCAAGCAATCCCATTTGGGGCATCAAGG 4500
 Qy 4501 GGGGGAGGATCTCATTTTCGCTTCCAAAGAAATGTGACGAGCTCGCGCAAGG 4560
 Db 4501 GGGGGAGGATCTCATTTTCGCTTCCAAAGAAATGTGACGAGCTCGCGCAAGG 4560
 Qy 4561 TGACAGGCTTCGAGCTGAAAGCTGTAGCATATTTACCGGGCTTGTGTGTGCTCATAC 4620
 Db 4561 TGACAGGCTTCGAGCTGAAAGCTGTAGCATATTTACCGGGCTTGTGTGTGCTCATAC 4620
 Qy 4621 CGGCTATTCGAGACGCTGCTGTGCTGGCAACAGACGCTCTAATGACGGGTTTACCGGCG 4680
 Db 4621 CGGCTATTCGAGACGCTGCTGTGCTGGCAACAGACGCTCTAATGACGGGTTTACCGGCG 4680
 Qy 4681 ATTTTGACTCAGTATGACTGCAATACATGTGTACCCAGACAGTCTGACTTCAGCTTGG 4740
 Db 4681 ATTTTGACTCAGTATGACTGCAATACATGTGTACCCAGACAGTCTGACTTCAGCTTGG 4740
 Qy 4741 ATCCCACTTACCATTCGAGACGACGACCGCTGCGGCTTGTGCTGCTGCGCAAC 4800
 Db 4741 ATCCCACTTACCATTCGAGACGACGACCGCTGCGGCTTGTGCTGCTGCGCAAC 4800
 Qy 4801 GGGAGGTAGAACTGGCAGGGGTAGGAGTGGCATCTACAGTTTGTGACTTCCAGGAGAAC 4860
 Db 4801 GGGAGGTAGAACTGGCAGGGGTAGGAGTGGCATCTACAGTTTGTGACTTCCAGGAGAAC 4860
 Qy 4861 GGGCTTCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
 Db 4861 GGGCTTCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
 Qy 4921 GGTATGAGCTCAGCGGCTGAGACCTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4980
 Db 4921 GGTATGAGCTCAGCGGCTGAGACCTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4980
 Qy 4981 GGTTCGCGCTGCGCAGGACCATCTGGAGTCTGGGAGGCTTACCTAATACACAG 5040
 Db 4981 GGTTCGCGCTGCGCAGGACCATCTGGAGTCTGGGAGGCTTACCTAATACACAG 5040
 Qy 5041 ACATGAGTCCGCTTCTGCTCCAGACTTAAACAGGCGAGGAGCAACTTTCTTACCTGCG 5100
 Db 5041 ACATGAGTCCGCTTCTGCTCCAGACTTAAACAGGCGAGGAGCAACTTTCTTACCTGCG 5100
 Qy 5101 TGCGCATATCAAGCTACAGTGTGCGGAGGCTCAAGTCTCACTCCATCTGCGGAGGAG 5160
 Db 5101 TGCGCATATCAAGCTACAGTGTGCGGAGGCTCAAGTCTCACTCCATCTGCGGAGGAG 5160
 Qy 5161 TGTGGAAGTGTCTCATACGCTGAAACCTTACCTGCAAGGCTCAAGTCTCACTCCATCTGCGGAGGAG 5220
 Db 5161 TGTGGAAGTGTCTCATACGCTGAAACCTTACCTGCAAGGCTCAAGTCTCACTCCATCTGCGGAGGAG 5220
 Qy 5221 GGTTCGAGCGGCTCAAAATGAGGTCTCTCCTCAGACACCCCATTAATAATCATCATGG 5280
 Db 5221 GGTTCGAGCGGCTCAAAATGAGGTCTCTCCTCAGACACCCCATTAATAATCATCATGG 5280
 Qy 5281 CATGCAATGCGGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5340
 Db 5281 CATGCAATGCGGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5340
 Qy 5341 TTTGAGGCTTGGCGGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5400
 Db 5341 TTTGAGGCTTGGCGGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5400
 Qy 5401 TCTTGTCCGGAAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5460
 Db 5401 TCTTGTCCGGAAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5460
 Qy 5461 AGATGGAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5520
 Db 5461 AGATGGAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5520

5521 ANTTCAAGCAAAAGCGCTCGGCTTGTTCACAAAGGCGCCACCAAGCAAGCGGAGCTGCTG 5580
 5521 AATTTTAAGCAAAAGCGCTCGGCTTGTTCACAAAGGCGCCACCAAGCAAGCGGAGCTGCTG 5580
 5581 CTCCTGGTGGAGTCCAAAGTGGCGAGCCCTTGAGACCTTCTGAGGCGGAAAGCAATGTGGA 5640
 5581 CTCCTGGTGGAGTCCAAAGTGGCGAGCCCTTGAGACCTTCTGAGGCGGAAAGCAATGTGGA 5640
 5641 ATTTCAATACCGGAATACAGTACTAGCAGGCTTATTCACATCTGCTGGGAAACCCCGCGA 5700
 5641 ATTTCAATACCGGAATACAGTACTAGCAGGCTTATTCACATCTGCTGGGAAACCCCGCGA 5700
 5701 TAGCATCATGTGGCATTTACAGCTTCTATCATAGCCCGCTCACACACCAAAACACCC 5760
 5701 TAGCATCATGTGGCATTTACAGCTTCTATCATAGCCCGCTCACACACCAAAACACCC 5760
 5761 TCTCTTTAATCATCTTGGGGGATGGGTGGTGTGCCAACTCGTCTTCCAGGCGTGGCT 5820
 5761 TCTCTTTAATCATCTTGGGGGATGGGTGGTGTGCCAACTCGTCTTCCAGGCGTGGCT 5820
 5821 CAGCTTTCTGGGCGCGGATCCCGGAGCGGCTGTGGCAGCATAGGCTTGGGAAG 5880
 5821 CAGCTTTCTGGGCGCGGATCCCGGAGCGGCTGTGGCAGCATAGGCTTGGGAAG 5880
 5881 TGCTCGTGGACATCTTGGCGGGCTATCGGGCAGGGGTAGCGGCGCACTCGTGGCCCTTA 5940
 5881 TGCTCGTGGACATCTTGGCGGGCTATCGGGCAGGGGTAGCGGCGCACTCGTGGCCCTTA 5940
 5941 AGGTCATAGCGGCGAGGTGCCCTCCACCGAGGACCTGGTCAACTTACTTCCCTGCCATCC 6000
 5941 AGGTCATAGCGGCGAGGTGCCCTCCACCGAGGACCTGGTCAACTTACTTCCCTGCCATCC 6000
 6001 TCTCTCTGTGGCTGGTCTGGGCGGCTGTGGCAGCAATATCTGCTGCGCAGCTGG 6060
 6001 TCTCTCTGTGGCTGGTCTGGGCGGCTGTGGCAGCAATATCTGCTGCGCAGCTGG 6060
 6061 GCCCGGAGAGGGGCTGTGAGTGGATGAACCGGCTGATAGCCTTCCGCGGGGTA 6120
 6061 GCCCGGAGAGGGGCTGTGAGTGGATGAACCGGCTGATAGCCTTCCGCGGGGTA 6120
 6121 ACCAGCTTCCCTACGCACTATGTGCTGAGAGGAGCGCTGACAGCTGTCTCATGAGA 6180
 6121 ACCAGCTTCCCTACGCACTATGTGCTGAGAGGAGCGCTGACAGCTGTCTCATGAGA 6180
 6181 TCTCTCTAGCTTACCATCTCATCACTGCTGAGCGGCTCCACAGTGGATTAATGAGG 6240
 6181 TCTCTCTAGCTTACCATCTCATCACTGCTGAGCGGCTCCACAGTGGATTAATGAGG 6240
 6241 ACTGCTCTACGCCATGTCCGGCTCGTGGCTAAGGGATGTTTGGGATTTGGATATGACAGG 6300
 6241 ACTGCTCTACGCCATGTCCGGCTCGTGGCTAAGGGATGTTTGGGATTTGGATATGACAGG 6300
 6301 TGTTCAGTCACTTAAGACCTGGCTCCAGTCCAACTCTGCGCGGCTTACCGGAGTCC 6360
 6301 TGTTCAGTCACTTAAGACCTGGCTCCAGTCCAACTCTGCGCGGCTTACCGGAGTCC 6360
 6361 CTCTTCTGTGTCAGCAAGCGGCTTACAGGAGTCTGGCGGGGACGCAATCATGCAAA 6420
 6361 CTCTTCTGTGTCAGCAAGCGGCTTACAGGAGTCTGGCGGGGACGCAATCATGCAAA 6420
 6421 CCACCTGCCCATCGGAGACAGATCGCGGACATGTCAAAAACGGTTTCCATGAGGATCG 6480
 6421 CCACCTGCCCATCGGAGACAGATCGCGGACATGTCAAAAACGGTTTCCATGAGGATCG 6480
 6481 TAGGGCTTGAACCTGACAGAAACAGTGGCAAGGAGTCTCCCATCAACGATACACCA 6540
 6481 TAGGGCTTGAACCTGACAGAAACAGTGGCAAGGAGTCTCCCATCAACGATACACCA 6540
 6541 CGGGACCTTGGCACCTTCCCGGCGCCAACTATTCCAGGGGCTTATGGCGGCTGGCTG 6600
 6541 CGGGACCTTGGCACCTTCCCGGCGCCAACTATTCCAGGGGCTTATGGCGGCTGGCTG 6600
 6601 CTGAGGAGTACGTGGAGTTACGCTGTGGGGGATTTCCACTACGTGACGGGCAATGACCA 6660

6601 CTGAGGAGTACGTGGAGTTACGCTGTGGGAGATTCCTACTACGTACGGGCATGACCA 6660
 6661 CTGAGGAGTACGTGGAGTTACGCTGTGGGAGATTCCTACTACGTACGGGCATGACCA 6660
 6661 CTGAGGAGTACGTGGAGTTACGCTGTGGGAGATTCCTACTACGTACGGGCATGACCA 6660
 6661 CTGAGGAGTACGTGGAGTTACGCTGTGGGAGATTCCTACTACGTACGGGCATGACCA 6660
 6721 GAGTGGGTTGACAGGTTACGCTCGGCGTGCACAACTCTCTCTACGGGAGGACGTACGT 6780
 6721 GAGTGGGTTGACAGGTTACGCTCGGCGTGCACAACTCTCTCTACGGGAGGACGTACGT 6780
 6781 TCCAGGTCGGGCTCAACCAATACCTTGGTGGTGGAGTCCCATGCGAGCCCGAACCG 6840
 6781 TCCAGGTCGGGCTCAACCAATACCTTGGTGGTGGAGTCCCATGCGAGCCCGAACCG 6840
 6841 ACCTAACAGTGTCTTCTTCCATGCTCACCGATCCCTCCACATTTACAGCAGAGACGGCTA 6900
 6841 ACCTAACAGTGTCTTCTTCCATGCTCACCGATCCCTCCACATTTACAGCAGAGACGGCTA 6900
 6901 AGCTTAGGCTGGCTAGAGGTTCTCCCTCTTTCAGCAGCTCATCAGTACGCACTTGT 6960
 6901 AGCTTAGGCTGGCTAGAGGTTCTCCCTCTTTCAGCAGCTCATCAGTACGCACTTGT 6960
 6961 CTGCGCTTCTTTGAAAGGCGACATGACATACCCATGACTCCCGGAGCGCTGACCTCA 7020
 6961 CTGCGCTTCTTTGAAAGGCGACATGACATACCCATGACTCCCGGAGCGCTGACCTCA 7020
 7021 TCCAGGCGCAACCTTCTTGGCGGAGGATGGGCGGAAACATCACTCGCTGGAGTCA 7080
 7021 TCCAGGCGCAACCTTCTTGGCGGAGGATGGGCGGAAACATCACTCGCTGGAGTCA 7080
 7081 AGAATAAGGTAGTAAATTTGGACTCTTTGAAACCGCTTTCAGCGGAGGGGATGAGGG 7140
 7081 AGAATAAGGTAGTAAATTTGGACTCTTTGAAACCGCTTTCAGCGGAGGGGATGAGGG 7140
 7141 AGATATCCGTCGCGGCGGAGATCCCGGAAATCCAGGAGTTCCTCCTCAGCGTTGCCCA 7200
 7141 AGATATCCGTCGCGGCGGAGATCCCGGAAATCCAGGAGTTCCTCCTCAGCGTTGCCCA 7200
 7201 TATGGGCAAGCGGACTCAATCTCTCACTGCTAGAGTCTCTGGAAGGACCCGCACTACG 7260
 7201 TATGGGCAAGCGGACTCAATCTCTCACTGCTAGAGTCTCTGGAAGGACCCGCACTACG 7260
 7261 TCCCTCGGCTGGTACAGGATGCGCACTGCTGCAAGGCTCTCCCAATACCACTC 7320
 7261 TCCCTCGGCTGGTACAGGATGCGCACTGCTGCAAGGCTCTCCCAATACCACTC 7320
 7321 CACGGAAGAGAGAGAGCGGTTGTCTGACAGATCCAAATGTCTTCTGCTTGGCGGAGC 7380
 7321 CACGGAAGAGAGAGAGCGGTTGTCTGACAGATCCAAATGTCTTCTGCTTGGCGGAGC 7380
 7381 TCGCCACTAAGACCTTCCGTTAGCTCGGATGCTCGGCGGTTGATAGCGGACCGGACCG 7440
 7381 TCGCCACTAAGACCTTCCGTTAGCTCGGATGCTCGGCGGTTGATAGCGGACCGGACCG 7440
 7441 CCTTCTGAGCTCGGCTCCGACAGCGTGCAGAGGATCCGAGCTTGAAGTCTGCTCTCT 7500
 7441 CCTTCTGAGCTCGGCTCCGACAGCGTGCAGAGGATCCGAGCTTGAAGTCTGCTCTCTCT 7500
 7501 CCATGCCCCCTTTGAAGGGGAGCGGCGGACCCCGATCTCAGCGACGGGCTTTGGTCTA 7560
 7501 CCATGCCCCCTTTGAAGGGGAGCGGCGGAGCCCGGATCTCAGCGACGGGCTTTGGTCTA 7560
 7561 CCGTGAAGTGAAGGCTAGTGAAGATGCTGCTGCTCAATGCTCAATGCTCAATGCTGAGCAG 7620
 7561 CCGTGAAGTGAAGGCTAGTGAAGATGCTGCTGCTCAATGCTCAATGCTCAATGCTGAGCAG 7620
 7621 GCGCCCTGATCAGCCATCGCTGCGGAGGAAAGTAAAGTCCCATCAACCCGTTGAGCA 7680
 7621 GCGCCCTGATCAGCCATCGCTGCGGAGGAAAGTAAAGTCCCATCAACCCGTTGAGCA 7680
 7681 ACTCTTCTGCTGCTCACACCAACATGCTGCTAGCGCACCAACATCCCGAGCGCAAGCTCC 7740

Db 7681 ACTCTTTGCTGCTCACCACAAACATGGTCTTAAGCCCAACACATCCCGCAGCGCAAGCCTCC 7740
Qy 7741 GGCAGAGAGAGTCACTTTTGACAGATTCGAGTCTCTGATGATCAATTACCAGGAGCTAC 7800
Db 7741 GGCAGAGAGAGTCACTTTTGACAGATTCGAGTCTCTGATGATCAATTACCAGGAGCTAC 7800
Qy 7801 TCAAGAGATGAAGGCGAAGGCGTCCACAGTTAAGGCTAAGCTTCTATCTATAGAGAGG 7860
Db 7801 TCAAGAGATGAAGGCGAAGGCGTCCACAGTTAAGGCTAAGCTTCTATCTATAGAGAGG 7860
Qy 7861 CTTGCAAGCTGACGCCCCCAATCGGCCAAATCCAAATTTGGCTATGCGGCAAGGAGCG 7920
Db 7861 CTTGCAAGCTGACGCCCCCAATCGGCCAAATCCAAATTTGGCTATGCGGCAAGGAGCG 7920
Qy 7921 TCCGAGACCTATCCAGAGGCGGCTTAAACACATCCGCTCCGCTCCGCTGCGGAGGACTTCGTCG 7980
Db 7921 TCCGAGACCTATCCAGAGGCGGCTTAAACACATCCGCTCCGCTCCGCTGCGGAGGACTTCGTCG 7980
Qy 7981 AAGACACTGAAAACACCAATTTGACACCACTCATGCGCAAAAGTGAAGTCTTCTGGGTCC 8040
Db 7981 AAGACACTGAAAACACCAATTTGACACCACTCATGCGCAAAAGTGAAGTCTTCTGGGTCC 8040
Qy 8041 AACAGAGAGGAGGCGGCAAGCAGCTCGCTTATTCGATTCCTCCAGACCTCGGAGTTC 8100
Db 8041 AACAGAGAGGAGGCGGCAAGCAGCTCGCTTATTCGATTCCTCCAGACCTCGGAGTTC 8100
Qy 8101 GTCTATCGAGAGAGTGGCCCTTTAGCAGTGGTCTCCACCTTCTCCAGGCGGTGATGG 8160
Db 8101 GTCTATCGAGAGAGTGGCCCTTTAGCAGTGGTCTCCACCTTCTCCAGGCGGTGATGG 8160
Qy 8161 GCTCCTCATACGAGATTCATTAATCTCCGCCAAGCAGCGGTCGAGTCTCCTGGTGAATACCT 8220
Db 8161 GCTCCTCATACGAGATTCATTAATCTCCGCCAAGCAGCGGTCGAGTCTCCTGGTGAATACCT 8220
Qy 8221 GGAATCAAGAGAGTGGCTTCTCATATGACACCGCGTGTGTTGACTCAACGG 8280
Db 8221 GGAATCAAGAGAGTGGCTTCTCATATGACACCGCGTGTGTTGACTCAACGG 8280
Qy 8281 TCACTGAGTGCATTCGTTGAGGAGTCAATTTACCAATGTGTGACTTGGCCCGCG 8340
Db 8281 TCACTGAGTGCATTCGTTGAGGAGTCAATTTACCAATGTGTGACTTGGCCCGCG 8340
Qy 8341 AGCCACAGAGGCGCATTAAGTTCGCTACAGAGCGGCTTTATCGGGGGTCCCTCGACTA 8400
Db 8341 AGCCACAGAGGCGCATTAAGTTCGCTACAGAGCGGCTTTATCGGGGGTCCCTCGACTA 8400
Qy 8401 ACTCAAGGCGAGACTGCGGTATTCGCGTTCGCGCGCAAGTGGCGTGTGACGACTA 8460
Db 8401 ACTCAAGGCGAGACTGCGGTATTCGCGTTCGCGCGCAAGTGGCGTGTGACGACTA 8460
Qy 8461 GCTGCGTAAATACCTCAGATGTTACTTTGAGGCGCACTGAGCGCTGTGCGAGTCAAGC 8520
Db 8461 GCTGCGTAAATACCTCAGATGTTACTTTGAGGCGCACTGAGCGCTGTGCGAGTCAAGC 8520
Qy 8521 TCCAGGACTGACGATGCTCGTGAAGCAGAGCGACCTTGTGTTATCTGTGAAGCGCG 8580
Db 8521 TCCAGGACTGACGATGCTCGTGAAGCAGAGCGACCTTGTGTTATCTGTGAAGCGCG 8580
Qy 8581 GAACCCAGGAGATGCGCGGCGCTTACAGCGTTCAGGAGCTATGACTAGGTATTCG 8640
Db 8581 GAACCCAGGAGATGCGCGGCGCTTACAGCGTTCAGGAGCTATGACTAGGTATTCG 8640
Qy 8641 CCCCCCGGGATCCGCCCCAACAGAAATACGACCTGGAGCTGATCAATCATGTTCT 8700
Db 8641 CCCCCCGGGATCCGCCCCAACAGAAATACGACCTGGAGCTGATCAATCATGTTCT 8700
Qy 8701 CCAATGTGTGAGTCCGCGCAGATCATCTGCGCAAAAGGTATACCTACCTACCGGTGACC 8760
Db 8701 CCAATGTGTGAGTCCGCGCAGATCATCTGCGCAAAAGGTATACCTACCTACCGGTGACC 8760
Qy 8761 CCACCAACCCCTTGGAGCGGCTGCGTGGGAGAGAGCTAGACACACTCCATCAACTCTT 8820
Db 8761 CCACCAACCCCTTGGAGCGGCTGCGTGGGAGAGAGCTAGACACACTCCATCAACTCTT 8820

Qy 8821 GCTAGGCAATATCATGTTATGCGCCACACCTATGSCAAGGATGATTTCTGATGACTC 8880
Db 8821 GCTAGGCAATATCATGTTATGCGCCACACCTATGSCAAGGATGATTTCTGATGACTC 8880
Qy 8881 ACTTTTTCCTCCATTCCTTCTAGCTCAAGAGCAACTTGAATAAGCCCTGATTTCTGATG 8940
Db 8881 ACTTTTTCCTCCATTCCTTCTAGCTCAAGAGCAACTTGAATAAGCCCTGATTTCTGATG 8940
Qy 8941 ACGGGCTTGTACTTCCATTGAGCCACTTGACCTACCTCAGATCAATTGAGACTTCCATG 9000
Db 8941 ACGGGCTTGTACTTCCATTGAGCCACTTGACCTACCTCAGATCAATTGAGACTTCCATG 9000
Qy 9001 GTCTTAGCCATTTACACTCCACAGTACTCTCCAGGTGAGATCAATAGGTTGGCTTCA 9060
Db 9001 GTCTTAGCCATTTACACTCCACAGTACTCTCCAGGTGAGATCAATAGGTTGGCTTCA 9060
Qy 9061 GCTCAGGAAACTTGGGTTACCACTTCCGAACTTGGAGACATCGGGCCAGAGTGTCC 9120
Db 9061 GCTCAGGAAACTTGGGTTACCACTTCCGAACTTGGAGACATCGGGCCAGAGTGTCC 9120
Qy 9121 GCGCTAAGTACTTCCAGGGGGAGGCGCCACTTGTGCGATACCTTCTTAACT 9180
Db 9121 GCGCTAAGTACTTCCAGGGGGAGGCGCCACTTGTGCGATACCTTCTTAACT 9180
Qy 9181 GGCAGTAAGGACCAAGCTTAAACTCACTCCCAATCCCGCGCGGTCCAGCTGGACTTGT 9240
Db 9181 GGCAGTAAGGACCAAGCTTAAACTCACTCCCAATCCCGCGCGGTCCAGCTGGACTTGT 9240
Qy 9241 CTGGCTGTGCTGCTGTGTTACAGGGGGAGACATATATCAGACCTGTCTCGTGCC 9300
Db 9241 CTGGCTGTGCTGCTGTGTTACAGGGGGAGACATATATCAGACCTGTCTCGTGCC 9300
Qy 9301 GACCCGCTGTTTCCGTTGCTGCTACTCTTCTAGGGTAGCATTTACTGTC 9360
Db 9301 GACCCGCTGTTTCCGTTGCTGCTACTCTTCTAGGGTAGCATTTACTGTC 9360
Qy 9361 TCCCAACCGATGAACGGGAGCTAACCTCAGGCTTAAGCCATTTCTGTTTTT 9420
Db 9361 TCCCAACCGATGAACGGGAGCTAACCTCAGGCTTAAGCCATTTCTGTTTTT 9420
Qy 9421 TT 9457
Db 9421 TT 9457

RESULT 7

HPCJ483 9448 bp RNA linear VRL 01-FEB-2000
Hepatitis C virus genome, complete sequence.
DEFINITION D13558 D01217
ACCESSION D13558.1 GI:221604
VERSION C protein; E protein; E2 protein; NS1 protein; NS2 protein; NS3 protein; NS4 protein; NS5 protein.
KEYWORDS Hepatitis C virus
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 9448)
Okamoto,H., Kojima,M., Okada,S., Yoshizawa,H., Iizuka,H., Tanaka,T., Muchmore,E.E., Peterson,D.A., Ito,Y. and Mishiro,S. Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability
Virology 190 (2), 894-899 (1992)
JOURNAL 92391112
MEDLINE 1325713
PUBMED 2 (bases 1 to 9448)
Okamoto,H.
AUTHORS Direct Submission
TITLE Submitted (17-OCT-1991) Hiroaki Okamoto, Jichi Medical School, Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi 329-04, Japan (E-mail:hokamoto@jichi.ac.jp, Tel:0285-44-2111(ex.3334), Fax:0285-44-1557);

Submitted (17-Oct-1991) to DDBJ by:

Hiroaki Okamoto
Immunology Division
Jichi Medical School
Kawachi-gun
Tochigi 329-04
Japan

Phone: 0285-44-2111 x3334

Fax: 0285-44-1557

Location/Qualifiers

1..9448

/organism="Hepatitis C virus"

/mol_type="genomic RNA"

/strain="HC-74"

/isolate="HC-74/83"

/db_xref="taxon:11103"

342..9374

/codon_start=1

/product="polyprotein precursor"

/protein_id="BA02756.1"

/db_xref="GI:221608"

/translation="MSINPKPQRTKNTNRPDQVFKPGGGQIVGGVLLPRGPRL
GVATKTSRSQPRMRQPIKARPEGRWAPQGPMPYNGNEGELWAGMLSPRG
SRPSWGPTDRRSRLGKVIDLTGCFADLMGYIPLVGLPGLGARALAHGVRELED
GVNATCNLPCCSFSIPLLALLSLATIPASAEVRNVSGIYHNTDCSNSIYVEAD
MIMHTGCPVCREDSSRCWALPTLAARNA SVPTTIRRHVDLILVGAFAFCSAMV
VGLDCSFLVLSQIFTPSRPHRTVDCNCSIVPGLSHGRMAWMMWSPITALV
SQLRIPQAVWVAGHWGLAGLAIYSWGVNAKVLIVALLFAGVDGTYTSGGA
SHTTSLASLSPASQRIQVNTVNGSHWNRNLCNDSLHGTGFLAALFYTHRPNS
GCPERVASCRPIDWAGQNGPIITVTEPSPDQRPQWHYAPRPGIIPASQVCGPVYC
PTSPVVGHTDTSQVPTSGWNETDMLNTRPPQWFGTGMWNTSTGFTKCGG
PCNIGVGHTLTCTPCDKRPEATYTKGSGPMLTPRCLVDYFVRLWHYPTCNF
SIFKRYVGVGVEHRLNAA CNWTEGPCNLEDRLDRSELSPLLSTTTEWQILPCAFTL
PALSTGLIHLQNIIVDQVLYGVGSAFVSFALKEWYIILALLADARVACGLMMLL
IAQAAELHNDVINAASVAGAGILSLVFFCAAWIYKRLAPCARAYFGWPLIL
LLALPPRAVDEMAASCGGAVLGLVLTSLPYKYVFLTRILWVLOQGITRYFVR
MQWVPLNVRGGSDAIILITCAVHELIDITKLLAILLGPLMLQAGITRYFVR
AQLIRACMLVRKVAGHYQMAFMKLGALITGVYNELPLRDWAGLRDLAVAE
PVPSDMEVETIKWAGDTAAGDILGLPVARSRGRI FLGPADSLGOGWRLIAPIT
A:SQGTGVLGCIITSLTRDKNQVEGVQVSTATQSFATCLNGVCTWTVHGAGX
TLAGKGPITQMYNDLVLGVQAPPGARSMTPTCGSSDLVLTBHAVI PVRRG
DSRGSLLSPRVSLVKGSGGGLPCPSGVVGVFRAAVCTRGSAKAVDFIPVSEMIT
MRSPVVDNSPVPAPPTQVAHLHPTGSGSTKVPAAYAAQYKVLNLPVSAATL
GFGVAMPKAGHIDINRTGVTITGPGITVYTKGLADGCGSGGAYDIIICDECHS
TDSITLIGLVQAEAGARLVVLTATPGSVTPHPNEIEIGLNNGEI PPKYK
ALPIEAKGSHLIFCHSKKCDBLAAKLGLGLNAVAYYRGLDVSIVPIGDVAVVA
TDLMTGTGDFDVIDNCTVOTVDPSLDPTITITITVTPQAVSRSRGTGSG
RSGLYFVTPGERSGMPDSSLCEYDAGCAWYELTPAETSVRLRAYLNTPGLPVQC
DHLFWESVFTGLSHIDAHFLSQTKQAGDNFPXLYVQATVCAQAQPPSPSWQWKC
LIRLPTLHGPTLLYRIGAVQNEVTLTHPTIKYIMACMSADLEVTTSTWLVGVLA
ALAAVCLTGVSVIYGRILISLGPAPVDPREVLYQEPDEMECASQLPYIEQGMQAR
QPKAIGLLATQAKQAAAPVSVESKWRALETFWAHMMNFI SGIQVLAGSLTPGN
PALASLMAFTASITSPLTQNTLILNGLHWVAGLAPPSPAASAFVAGIAGAAGVSI
GLGLVLDILAGYAGVAGALVAFKVMGSEVPSTEDLVNLLPAILSPGALVGVVCA
ILRRHVGEGAVQMMNRLIAPSRGNHVSPTHVYVESDAARVTQILSSLTITQLLK
RLHWINDECTPCSGSDPTDVLWDMCTVLDFTKWLQSLKILPRLEPFLSCORGYK
GWRGDMTQTCPCGAOITAGHVXGSMRI VGPRTCSNTHGTFPINAYTTGCTPSP
APNYSRLARWAEVSVTRGVDFHVTGMTDNNKCPQCPVAPPEPTVEVDVGLLR
YAPACKLLREDVAFQVGLNQVLVGSQLPCEPEPDVTLTSMIDPSHITAEAKRL
ARGSPSLASSASQLSAPSKATCTTHHSDPADLIEANLWRQMGENTIRVSEN
KVIILDFSEPLHAEGBREISVAABILRKSXFPSPALPFWARPYNPMLLSWDDPY
VPVVGCPPLPTKAPPIPPRRKRTVLTESNVSSALAEATKTFSSGSSAVDSMS
ATALLPOASDDKGSVDSESYSPMPLEGPGLDSDGSMSTVSEBESDEDVCSMS
VMTGALLTPCAESKULPINLSLLRHNMVATTSRSASLRQKVTFDRLVOLD
DHYRDLKEMKAKASTYKAKLLISEAKLITPPHSAKSPFGYKADVNLSRAVNH
RSVBELEDTEPTDITTIMAKNEVFCVOPERGRKAPKLI VFDLGRVRCVKMALYD
VYSTLPQAVNGSYQFQSPKORVFLNNTWKSQCPMPGFSYDTRCFDSTVTESDIRV
BESIYQCCDLAEARQAIRSLTERLYIGGLTNKQNGCYRCRASGLVTSNCTIL
TCVLKATAACRAKQDCVMLVNGDLVVICESAGTQEDAAALRYETAMTRYSPGP
DPQPEYDLRLITSGSNYSVAHDASGRVYVLTTRDPTIPLARAWEHTARHTPVNSWL
GNIMYAPALWARMILMTHFSLIAQEOLEQALDCQIYGACYSIEPLDLPOLIERLH
GLSAFSLSHSYSGEINRVASCLRKUGVPLPRVHRFASVRAKLISQGRATCYRIL
FNNAVETLKLFTPIFAASQLDLSGWFVAGYSGGDIYHLSUSRARPRFWFLICLLDLSVGV

GIYLLPNR"
342..914
/product="C protein"
915..1490
/product="E protein"
1491..2528
/product="NS1/E2 protein"
2529..3359
/product="NS2 protein"
3360..5186
/product="NS3 protein"
5187..6380
/product="NS4 protein"
6381..9371
/product="NS5 protein"
6005..6104
/note="possible stem-loop structure in NS4 sequence"
9413..9448
/note="1-stretch of 36 bp"

ORIGIN

Query Match 95.8%; Score 9190.4; DB 14; Length 9448;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 9287; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 1 GCACGCCCCCTGATGGGGGCGGACACTCCACCATAGATCACTCCCTCTGAGGAATCTATG 60
DB 1 GCACGCCCCCGATGGGGGCGGACACTCCACCATAGATCACTCCCTCTGAGGAATCTATG 60
QY 61 TCTTCCGCGAGAAGCGTCTAGCCATGGGTAGTATGATGTCGTGACGCTCCAGGAC 120
DB 61 TCTTCCGCGAGAAGCGTCTAGCCATGGGTAGTATGATGTCGTGACGCTCCAGGAC 120
QY 121 CCCCCCTCCCGGAGAGCCATAGTGTCTCGGAAACCGGTGAGTACACCGGAATTCGCCAG 180
DB 121 CCCCCCTCCCGGAGAGCCATAGTGTCTCGGAAACCGGTGAGTACACCGGAATTCGCCAG 180
QY 181 GAGACCGGGTCTCTTCTTGGATCAACCCCTCAAATGCTTGGAGATTTCGGGCTGCCCC 240
DB 181 GAGACCGGGTCTCTTCTTGGATCAACCCCTCAAATGCTTGGAGATTTCGGGCTGCCCC 240
QY 241 GCGAGCTGTAGCCAGTAGTGTGGTTCGGAAGAGCCCTTGTGTACTTGCCTGATAGG 300
DB 241 GCGAGCTGTAGCCAGTAGTGTGGTTCGGAAGAGCCCTTGTGTACTTGCCTGATAGG 300
QY 301 GTGCTTCGAGTGCCTCGGAGGTCTCGTAGACCGTGCACATGAGCAGCAATCTTAAAC 360
DB 301 GTGCTTCGAGTGCCTCGGAGGTCTCGTAGACCGTGCACATGAGCAGCAATCTTAAAC 360
QY 361 CTCAGAGAAAACCAACGTAACCAACCGCCGCCACAGGAGCTCAAGTCCCGGGCG 420
DB 361 CTCAGAGAAAACCAACGTAACCAACCGCCGCCACAGGAGCTCAAGTCCCGGGCG 420
QY 421 GTGCTCAGATCGTTGGTGGAGTTTACCTGTTCGCGCGCAGGGGCCCCAGGTTGGGTGTC 480
DB 421 GTGCTCAGATCGTTGGTGGAGTTTACCTGTTCGCGCGCAGGGGCCCCAGGTTGGGTGTC 480
QY 481 GCGGACTTAGGAAGCTTCCGAGCGGTCCGAACTCGTGTGAAGCGGACCACTATCCCA 540
DB 481 GCGGACTTAGGAAGCTTCCGAGCGGTCCGAACTCGTGTGAAGCGGACCACTATCCCA 540
QY 541 AGGCTCGCGACCCGAGGGAGGCTTGGGTGAGCGGGGTGAGTCCCTTGGCCCTCTATG 600
DB 541 AGGCTCGCGACCCGAGGGAGGCTTGGGTGAGCGGGGTGAGTCCCTTGGCCCTCTATG 600
QY 601 GCAATGAGGGCTGGGGTGGGAGGATGGTCTCTGTCAACCCCGCGGTCCCGGCTAGTT 660
DB 601 GCAATGAGGGCTGGGGTGGGAGGATGGTCTCTGTCAACCCCGCGGTCCCGGCTAGTT 660
QY 661 GGGGCCCCACGAGCCCCCGGGGTAGGTGCGCTAACTTGGTAAAGTCAATCATACCCCTTA 720
DB 661 GGGGCCCCACGAGCCCCCGGGGTAGGTGCGCTAACTTGGTAAAGTCAATCATACCCCTTA 720

721 CATGGGGCTTCGGGATCTCATGGGTACATTCGGCTCGTGGGCGCCCGCTAGGGGCG 780
721 CATCGGGCTTCGGGATCTCATGGGTATATTCGGCTCGTGGGCGCCCGCTAGGGGCG 780
781 CTGCCAGGCTTCGGACACGGGTCTCGGGTCTTGGAGACGGGCTGAATATGCAACAG 840
781 CTGCCAGGCTTCGGACACGGGTCTCGGGTCTTGGAGACGGGCTGAATATGCAACAG 840
841 GGAATTCGCCGGTTCCTTCTCTATCTTCCTCTTGGCTCTGCTCTCTCTGTTGACCA 900
841 GGAATTCGCCGGTTCCTTCTCTATCTTCCTCTTGGCTCTGCTCTCTCTGTTGACCA 900
901 TCCAGCTTCGGCTTATGAGAGTGGCAACGTCTCGGATATACCATGTCAGGAACGACT 960
901 TCCAGCTTCGGCTTATGAGAGTGGCAACGTCTCGGATATACCATGTCAGGAACGACT 960
961 GCTCCAACTCAAGCATTCGTATGAGGACGGGACGTGATCATGCACTATCCCGGGTGG 1020
961 GCTCCAACTCAAGCATTCGTATGAGGACGGGACGTGATCATGCACTATCCCGGGTGG 1020
1021 TGGCTGTGTTTACGAGGTAAAGCTTCGGTTCGTGGTAGGCTCACTCCCAAGCTCG 1080
1021 TGGCTGTGTTTACGAGGTAAAGCTTCGGTTCGTGGTAGGCTCACTCCCAAGCTCG 1080
1081 CGGCCAGGATCCAGGCTCCCACTACGACATACGACGCACTGCTGCTGCTGCTG 1140
1081 CGGCCAGGATCCAGGCTCCCACTACGACATACGACGCACTGCTGCTGCTGCTG 1140
1141 GGAGGCTGCTTTCCTGCTCCCTATGACGTGGGGATCTCTGGGATCTATTTCTCTG 1200
1141 GGAGGCTGCTTTCCTGCTCCCTATGACGTGGGGATCTCTGGGATCTATTTCTCTG 1200
1201 TCTCCAGCTGTTACCTTTCGGCTTCGGCTTCGGCTTCGGCTTCGGCTTCGGCTTC 1260
1201 TCTCCAGCTGTTACCTTTCGGCTTCGGCTTCGGCTTCGGCTTCGGCTTCGGCTTC 1260
1261 CAATCTATCCGGGCTATGACGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 1320
1261 CAATCTATCCGGGCTATGACGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 1320
1321 CACCTACACAGCTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTC 1380
1321 CACCTACACAGCTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTC 1380
1381 TGGTGGGGGGCCCACTGGGGAGTCTGGGGAGTCTGGGGAGTCTGGGGAGTCTGGGGAG 1440
1381 TGGTGGGGGGCCCACTGGGGAGTCTGGGGAGTCTGGGGAGTCTGGGGAGTCTGGGGAG 1440
1441 ACTGGGCTAAGGTTCTGATTTGGGGTACTCTTGGGGGCTTGGGGGAGTCTGGGGAG 1500
1441 ACTGGGCTAAGGTTCTGATTTGGGGTACTCTTGGGGGCTTGGGGGAGTCTGGGGAG 1500
1501 CGACGGGAGGTGGCGGACCAACCACTTCGGGTTCAGTTCGGGTTCAGTTCGGGTTCAG 1560
1501 CGTGGGGGGGGCGGCGGACCAACCACTTCAGGCTTCGGGTTCAGTTCGGGTTCAGG 1560
1561 CGTCTCAGAAATCAGCTTGTGAATCAACGCGAGCTGGCAGTCAACAGAGTGGCC 1620
1561 CGTCTCAGAAATCAGCTTGTGAATCAACGCGAGCTGGCAGTCAACAGAGTGGCC 1620
1621 TAAATTCGAATGACTCCCTCCAACTGGGTCTTTTGGCGGCTGTTTATCGCACCAAGT 1680
1621 TAAATTCGAATGACTCCCTCCAACTGGGTCTTTTGGCGGCTGTTTATCGCACCAAGT 1680
1681 TCAACTCGTTCGGGTTCGGGAGCGGACGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1740
1681 TCAACTCGTTCGGGTTCGGGAGCGGACGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1740
1741 AGGGGTGGGGCCCCATCACTATCTAAGCTAACAGCTCGGATCAGAGCCCTTATGCT 1800
1741 AGGGATGGGGCCCCATCACTATCTAAGCTAACAGCTCGGATCAGAGCCCTTATGCT 1800
1801 GGCATTTACGGCCCTCGACCGTGTGGTCTGTAACCGGCTCGCAGGTGTGTGGTCCAGTGT 1860

1801 GGCATTTACGGCCCTCGACCGTGTGGTCTGTAACCGGCTCGCAGGTGTGTGGTCCAGTGT 1860
1861 ATTGTTTCAACCCAGGCTGTTGGTGGGAGCACCGATGTTCCGGGTTCCTTACCT 1920
1861 ATTGTTTCAACCCAGGCTGTTGGTGGGAGCACCGATGTTCCGGGTTCCTTACCT 1920
1921 ATAGCTGGGGGAGATGAGACAGAGCTGATGCTCTCAACAAACAGCTTCGCCCAACAG 1980
1921 ATAGCTGGGGGAGATGAGACAGAGCTGATGCTCTCAACAAACAGCTTCGCCCAACAG 1980
1981 GCAACTGGTTCGGCTTATCATGATCAATAGTACTGGTTCATTAAGAGTGGGGAGGC 2040
1981 GCAACTGGTTCGGCTTATCATGATCAATAGTACTGGTTCATTAAGAGTGGGGAGGC 2040
2041 CCCCGTGAATACGCGGGGTCGGTAACGCAACCTTGATCTGCCCAACAGGAGTGTCTTC 2100
2041 CCCCGTGAATACGCGGGGTCGGTAACGCAACCTTGATCTGCCCAACAGGAGTGTCTTC 2100
2101 GGAAGCACCCCGAGGCTTACTTACACAAAATGTGGCTCGGGGCTCGGTTGACACCTAGGT 2160
2101 GGAAGCACCCCGAGGCTTACTTACACAAAATGTGGCTCGGGGCTCGGTTGACACCTAGGT 2160
2161 GCCTAGTAGACTACCCATACAGGCTCTGGCCTACCCCTGCACTTCAATTTTCACTCT 2220
2161 GCCTAGTAGACTACCCATACAGGCTCTGGCCTACCCCTGCACTTCAATTTTCACTCT 2220
2221 TTAAGGTTAGATGATGTTGGGGGCTGGAGCAACGCTCAATGCGGACGCAATTTGGA 2280
2221 TTAAGGTTAGATGATGTTGGGGGCTGGAGCAACGCTCAATGCGGACGCAATTTGGA 2280
2281 CTGAGGAGAGCGCTTAACTTGGAGACAGGATAGGTGAGAACTCAGCCCGCTGTCTG 2340
2281 CTGAGGAGAGCGCTTAACTTGGAGACAGGATAGGTGAGAACTCAGCCCGCTGTCTG 2340
2341 TGTCTACACAGAGTGCAGATACCTGCTGCTGCTTTCACACCTCACCCTTTTATCCA 2400
2341 TGTCTACACAGAGTGCAGATACCTGCTGCTGCTTTCACACCTCACCCTTTTATCCA 2400
2401 CTGGTTTGTATCCATCTCCATCAGAAACATCGTGGACGTCGCAATACCTGTAAGGTGAGGT 2460
2401 CTGGTTTGTATCCATCTCCATCAGAAACATCGTGGACGTCGCAATACCTGTAAGGTGAGGT 2460
2461 CAGCGTTTGTCTCTCTTTGCAATCGGAGTACATCTGCTGCTTTCCTTCTCTCTG 2520
2461 CAGCGTTTGTCTCTCTTTGCAATCGGAGTACATCTGCTGCTTTCCTTCTCTCTG 2520
2521 CAGACGCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
2521 CAGACGCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
2581 CTTTAGAGAACTTGGTGGTCTCAATGCGGCTCGTGGCGGAGCGCATGCTATTTCT 2640
2581 CTTTAGAGAACTTGGTGGTCTCAATGCGGCTCGTGGCGGAGCGCATGCTATTTCT 2640
2641 CTTTCTTGTGCTTCT 2700
2641 CTTTCTTGTGCTTCT 2700
2701 CGTATGCTTTTATGGCGTATGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
2701 CGTATGCTTTTATGGCGTATGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
2761 CTTAGCGCTTGGACGGGAGATGGTGCATCGTGGGGGCTGGCTTCTTGTAGTCTG 2820
2761 CTTAGCGCTTGGACGGGAGATGGTGCATCGTGGGGGCTGGCTTCTTGTAGTCTG 2820
2821 TATTTCTGACCTTGTACCATCTACAAAGTCTTCTCCTAGGCTCATATGTTGTTTAC 2880
2821 TATTTCTGACCTTGTACCATCTACAAAGTCTTCTCCTAGGCTCATATGTTGTTTAC 2880
2881 AATACTTTATACAGAGCGGAGCGCATGCAAGTGTGGGTGCTGCTGCTGCTGCTGCTGCT 2940

Db 2881 AATACCTTTATACACAGAGCCGAGGCGCACATGACAGGTGTGGTCCGCCCTCAACGTTTC 2940
Qy 2941 GGGAGAGCCGCGATGCCATCATCTCTCTACGTGTGGGTTCATCCAGAGTTAATTTTTC 3000
Db 2941 GGGAGAGCCGCGATGCCATCATCTCTCTACGTGTGGGTTCATCCAGAGTTAATTTTTC 3000
Qy 3001 ACATCACCACAACTCCTGTGCGCATATCTCGCGCGCTCATGTGTCTCCAGCTGGCATAA 3060
Db 3001 ACATCACCACAACTCCTGTGCGCATATCTCGCGCGCTCATGTGTCTCCAGCTGGCATAA 3060
Qy 3061 CGAGAGTCCGCTACTCTGTGCGCGCTCAAGGGCTCATTCGTFGCATGATGTTAGTGCAG 3120
Db 3061 CGAGAGTCCGCTACTCTGTGCGCGCTCAAGGGCTCATTCGTFGCATGATGTTAGTGCAG 3120
Qy 3121 AAGTCGCGGGGCTCATTTAGTCCAAATGCTTCATGAAGTGGCGCGCTGACAGGTA 3180
Db 3121 AAGTCGCGGGGCTCATTTAGTCCAAATGCTTCATGAAGTGGCGCGCTGACAGGTA 3180
Qy 3181 CGTAGTGTATAACCATCTTACCCACTCGGGACTGGGCCACCGCGGCTTACGAGACC 3240
Db 3181 CGTAGTGTATAACCATCTTACTCTCATGCGGACTGGGCCACCGCGGCTTACGAGACC 3240
Qy 3241 TTGCGGTGCGGTAGAGCCCTGCTCTCTCCGCCATGAGACCAAGGTTCATCACTGGG 3300
Db 3241 TTGCGGTGCGGTAGAGCCCTGCTCTCTCCGCCATGAGACCAAGGTTCATCACTGGG 3300
Qy 3301 GAGCAGACACCTGCTGCTGGGACATCATCTTGGGTCTACCCGCTCTCCGCCGAGAGG 3360
Db 3301 GAGCAGACACCTGCTGCTGGGACATCATCTTGGGTCTACCCGCTCTCCGCCGAGAGG 3360
Qy 3361 GGAAGGAGATATTTTGGGACCGGCTGATGCTCTGGAAGGGCAAGGTTGGGACTCTCTT 3420
Db 3361 GGAAGGAGATATTTTGGGACCGGCTGATGCTCTGGAAGGGCAAGGTTGGGACTCTCTT 3420
Qy 3421 GCGCCATCAAGCCCTACTCCCAACAAACGCGGGCGGTACTTGGTGTGATCATCACTAGCC 3480
Db 3421 GCGCCATCAAGCCCTACTCCCAACAAACGCGGGCGGTACTTGGTGTGATCATCACTAGCC 3480
Qy 3481 TCACAGGCGGGACACAGACAGGTTCGAAGGGAGGTTCAAGTGGTTTCAACGCAACAC 3540
Db 3481 TCACAGGCGGGACACAGACAGGTTCGAAGGGAGGTTCAAGTGGTTTCCACCGCAACAC 3540
Qy 3541 AATCTTTCTCTGCGACCTGCATCAACGCGGTGTGCTGACATGCTTACCATGCGCTGGCT 3600
Db 3541 AATCTTTCTCTGCGACCTGCATCAACGCGGTGTGCTGACATGCTTACCATGCGCTGGCT 3600
Qy 3601 GBAAGACCTACCGGTTCBAAGGTTCBAATCACCCAAATGTACACCAATGTAGACCTGG 3660
Db 3601 GBAAGACCTACCGGTTCBAAGGTTCBAATCACCCAAATGTAGACCTGG 3660
Qy 3661 ACCTGCTGGCTGGCAGGCGCCCGCGGGCGCTCCATGACACCATGACATGACCTGGCA 3720
Db 3661 ACCTGCTGGCTGGCAGGCGCCCGCGGGCGCTCCATGACACCATGACATGACCTGGCA 3720
Qy 3721 GCTGGAACCTTTACTTGTGTACAGACATGCTGATGATGCTGATGCTGATGCTGATGCTG 3780
Db 3721 GCTGGAACCTTTACTTGTGTACAGACATGCTGATGATGCTGATGCTGATGCTGATGCTG 3780
Qy 3781 ACAGCAGGGAGTCTACTCTCCCCAGGCGCTCTCTACCTGAAAGGCTCTCTGGGTG 3840
Db 3781 ACAGCAGGGAGTCTACTCTCCCCAGGCGCTCTCTACCTGAAAGGCTCTCTGGGTG 3840
Qy 3841 GTCCATTGTTTTCCTTCCGCGCAGCTGCTGGGCGCTCTTCCGGGCTGCTGTGTGACACC 3900
Db 3841 GTCCATTGTTTTCCTTCCGCGCAGCTGCTGGGCGCTCTTCCGGGCTGCTGTGTGACACC 3900
Qy 3901 GGGGGTCCGAGAGCGGTGACTTCATACCCGTTGAGTCTATGGAATTAATATGCGGT 3960
Db 3901 GGGGGTCCGAGAGCGGTGACTTCATACCCGTTGAGTCTATGGAATTAATATGCGGT 3960
Qy 3961 CTCGCGTCTTCACAGACACTCAACCCCGCGCTGTACCGCAGACATTCACAGTGGCAC 4020
Db 3961 CTCGCGTCTTCACAGACACTCAACCCCGCGCTGTACCGCAGACATTCACAGTGGCAC 4020

Qy 4021 ATCTGCACGCTCTACTGTCAGCGGCAAGAGACCAAAAGTGCGGTGTGCTATGACGCC 4080
Db 4021 ATCTGCACGCTCTACTGTCAGCGGCAAGAGACCAAAAGTGCGGTGTGCTATGACGCC 4080
Qy 4081 AAGGTPACAAAGTGTCTGTCGACCCCTGCTGGTTCGCCCCACCTTAGGTTCGGGCT 4140
Db 4081 AAGGTPACAAAGTGTCTGTCGACCCCTGCTGGTTCGCCCCACCTTAGGTTCGGGCT 4140
Qy 4141 ATATGTCCAGGACACAGGTATGACCCCTAAATCAGAACTGGGTAAAGACATTAACA 4200
Db 4141 ATATGTCCAGGACACAGGTATGACCCCTAAATCAGAACTGGGTAAAGACATTAACA 4200
Qy 4201 CGGGCGCTCCATTACGTACTCCACTATGGCAAGTTCCTTCCCGACGGTGTGCTG 4260
Db 4201 CGGGCGCTCCATTACGTACTCCACTATGGCAAGTTCCTTCCCGACGGTGTGCTG 4260
Qy 4261 GGGCGCTATGACATCATATATGTGATGATGCGCACTCAACTGACTGACTACCATCT 4320
Db 4261 GGGCGCTATGACATCATATATGTGATGATGCGCACTCAACTGACTGACTACCATCT 4320
Qy 4321 TGGGCATCGGCACAGTCTCGACCAAGCGAGACGGCTGGAGCGCGCTCGTCTGCTCG 4380
Db 4321 TGGGCATCGGCACAGTCTCGACCAAGCGAGACGGCTGGAGCGCGCTCGTCTGCTCG 4380
Qy 4381 CCACCGCTACACCTCCGGGATCGGTACCGTGCCAACCCCAATATCGAGGAATAGGCC 4440
Db 4381 CCACCGCTACACCTCCGGGATCGGTACCGTGCCAACCCCAATATCGAGGAATAGGCC 4440
Qy 4441 TGTCACCAATGAGAGATCCCTCTTATGGCAAGCCATCCCATTTGAGGCCATCAAG 4500
Db 4441 TGTCACCAATGAGAGATCCCTCTTATGGCAAGCCATCCCATTTGAGGCCATCAAG 4500
Qy 4501 GGGGAGGCTCTCATTTCTGCAATTCAGAGAAATGTGACGAGCTCCCGCAAGC 4560
Db 4501 GGGGAGGCTCTCATTTCTGCAATTCAGAGAAATGTGACGAGCTCCCGCAAGC 4560
Qy 4561 TGACAGCCCTCGGACTCAATGCTGTAGCAATTTACCGGGCTTGTGCTGCTATAC 4620
Db 4561 TGACAGCCCTCGGACTCAATGCTGTAGCAATTTACCGGGCTTGTGCTGCTATAC 4620
Qy 4621 GCGCTATCGAGAGCTGTGTGTGTGCAACAGAGCTCTAATGACGGGTTTACCGGG 4680
Db 4621 GCGCTATCGAGAGCTGTGTGTGTGCAACAGAGCTCTAATGACGGGTTTACCGGG 4680
Qy 4681 ATTTTGTAGTCACTGATGCAATGATGTGTACCCAGACAGTCCGACTTCAGCTTGG 4740
Db 4681 ATTTTGTAGTCACTGATGCAATGATGTGTACCCAGACAGTCCGACTTCAGCTTGG 4740
Qy 4741 ATCCACCTTCCATTTGAGACGACGACCGTGTGCGGCAAGACGGGTTCGGCTCGCAC 4800
Db 4741 ATCCACCTTCCATTTGAGACGACGACCGTGTGCGGCAAGACGGGTTCGGCTCGCAC 4800
Qy 4801 GGGAGGTAGAACTGGCAGGGTAGGAGTGGCATCTACAGGTTTGTGACTCCAGAGAAC 4860
Db 4801 GGGAGGTAGAACTGGCAGGGTAGGAGTGGCATCTACAGGTTTGTGACTCCAGAGAAC 4860
Qy 4861 GGGCTCGGGCATGTTGCAATCTTCTGCTGTGTGAGTGTCTATGACGCGGCTGTGCTT 4920
Db 4861 GGGCTCGGGCATGTTGCAATCTTCTGCTGTGTGAGTGTCTATGACGCGGCTGTGCTT 4920
Qy 4921 GGTATGAGCTCACGCCCTGAGACCTCGGTAGGTTGCGGGCTTACTTAATACACAG 4980
Db 4921 GGTATGAGCTCACGCCCTGAGACCTCGGTAGGTTGCGGGCTTACTTAATACACAG 4980
Qy 4981 GGTTCGCCGTCTGCGAGACCATCTGGAGTTCTGGAGAGCGTCTTACAGGCTCTCCC 5040
Db 4981 GGTTCGCCGTCTGCGAGACCATCTGGAGTTCTGGAGAGCGTCTTACAGGCTCTCCC 5040
Qy 5041 ACATAGATGCCACTTCTCTATCCAGACTAAACAGCAGGAGACACTTCTTCTTACCTGG 5100
Db 5041 ACATAGATGCCACTTCTCTATCCAGACTAAACAGCAGGAGACACTTCTTCTTACCTGG 5100

RESULT 8
 AF054255
 LOCUS
 DEFINITION
 Hepatitis C virus HC-J4, pCV-J416 polyprotein gene, partial cds.
 ACCESSION
 AF054255
 VERSION
 AF054255.1 GI:3098647
 KEYWORDS
 Hepatitis C virus
 ORGANISM
 Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 REFERENCE
 1 (bases 1 to 8780)
 Yanagi, M., St. Claire, M., Shapiro, M., Emerson, S. U., Purcell, R. H. and
 Bukh, J.
 TITLE
 Transcripts of a chimeric cDNA clone of hepatitis C virus genotype
 1b are infectious in vivo
 JOURNAL
 Virology 244 (1), 161-172 (1998)
 MEDLINE
 98240944
 PUBMED
 9581788
 REFERENCE
 2 (bases 1 to 8780)
 Yanagi, M. and Bukh, J.
 AUTHORS
 Direct Submission
 TITLE
 Submitted (18-MAR-1998) Hepatitis Viruses Section, LID, NIAID, NIH,
 JOURNAL
 7 Center Dr., Building 7, Room 201, Bethesda, MD 20892, USA
 FEATURES
 source
 1..8780
 /organism="Hepatitis C virus"
 /mol_type="genomic RNA"
 /strain="HC-J4"
 /isolate="pCV-J416"
 /db_xref="taxon:11103"
 187..>8780
 /codon_start=1
 /product="polyprotein"
 /protein_id="AAC15729.1"
 /db_xref="GI:3098648"
 /translation="MSTNPKPQKTKENTNRPODKVFPKGGQIVGGVILLPRGPRL
 GVRATRKASQPRGRQPIKARRPGRNAWQFDPWELNYGNEGLWAGWLLSPRG
 SRPSWGTDRRRSRGKVIDTTCGFADMLGKIPVLPAGLPAGALAHGVRLVED
 GVNATGLPGCSFIFLLALLSLTIPASAYEVNVSIGYHVNDCSNSSIVVEAD
 VIMETPGCPVCGNSRSRWALPTLAPNAPSPTTIRRHVDLVLGTAAFCSAM
 VDLGGLIFLVSLFTSPRHRTVQDCNCSIPYGHVSGHRMAMDMWSPPTALV
 SOLLRIPQAVNVAGWAGVLAGLVYVMGNWAKVLI VALLFAGVDGTHHTGRVA
 GHTSGTSLSFSSGASOKIOLAVNTGSMHNRCLNDCNSLOQFPAALFYAHKFNSS
 CQPMASCRPIIDFPAQWGEIITTKENSDQRYCNHYARPCGVTPASQVCPVTC
 FTPSPVVTGIDRSQVPTYSWGENETDMLNLTNPQGNWFGCTWNSGTFTKTCG
 PPNIGVGNRLICPTDCPRKPEATYKCGSGFWLTPLCLVDYFVRLMHPCTINF
 SIFKVMYGVGEHLNAAACNWTGRGRCNLRDRRSELSPLLLSTTETWQLPCAFTNL
 PALSTGLIHLHONTVDVQYLYGVGSFAVSFAIKWEYILLFLLLADARVACILWMMLL
 IAQAEALENAVLNAAVSAGHGLSPLVFPCCAWYIKGLRPAQAVAFYGVWPLLL
 LLLALPPRAYALDREMAASCGAVLVGLVLTLSFYKYVFLTRLIWMLQVILITAEAH
 MQWVPLPLNVRGGKDAIILLCAVHPELIFDITKLLALLGLFLMWLQAGITRVFVFR
 AQGLIRACMLYKRVAGHYVQWFMKLGALGTXYVYNHLTELRDWAHAGLRDLAVAE
 PVFVSAMETKVITWGAATAAGDIIILGLPVARRGKEIFLGPADSLGQGWRLIAPIT
 AVSOOTRGVLCIITSLTRKNOVEGVVSTATQSFALGTCINGVCTWTVYHAGSK
 TLAAPKGPITOWNVLDLWQCAPPCARSWTPCSGSDLYLVTRHADVI PVRRRG
 DSRGLLSFPRVSVHLKSGSGPLPCPSGHVVGSPFAAVCTRGVAKAVDFI IVESEMET
 MRSFVTDNSFPFAPVQTFQVHLHAPTGSGKSTKVPAYAAQKGLVLNPSVAATL
 GFQYMSAHGINDIRNTGVTITGGITATYSTYKFLADGCGSGAYDIIICDECHS
 TDSATLIGTIVLDQAEAGARLVLTATPPGVSVPVPHNIIEIGLSNNGEIPFYK
 AIPBAIKGRLHIFCHSKCKDELAATLGLNNAVAYYRGLDVSVPPIGDVVVVA
 TDLMTGTGPDVSDVNDCTVOTVDFSLDPTFTIEITVTPQDASVRSRRGTRGR
 RSGIYRVPGRSPGMPDSVLCYCYDAGCANWELITAEYSVRLRAYLNTGLPVQ
 DHEBESWFTGLTHIDAHFSQTKQAGNDFYLVAYQATVCARAAQWPPSDQMWK
 LTRKPLHGTTPGLYRIGAVQNEVILTHPIKYMCMASADLEVVTWVLVGLVLA
 ALAAATLGTSGVIVGRIIILSGKPAVEDREVLYQFDEMBECASQLPFIIEQGMOLAE
 QPKALGLLOTATQOAAAPAVESKWRALFTFWAKHWNFIQSIOVLAGLSTLPNG
 PALASIMAFASITSLPTONTLLFNILGMSVAQIAPPSAASAFVAGIAGVAGSI
 GLGKVLIDLAGVAGVAGLVAFKVMGSEVPSTEDLVNLLPAILSPAGLVGVVCAI
 ILRRHVPGEGAVQWNRNALVAFASRGNESVPHYVPSDAAARVTQIILSLITITQLK
 RLHOWINEDCSTPCSGSLNDRYDWICTVLDFTDQLQSLPLRPLGVPVFFLSCRGYK
 GVMRGDINQITTCQGAQTAGEVKNKSNRIIVGPRTCNSNTWHGTFPINAITYTGPCTPSP

ORIGIN

Query Match 91.5%; Score 8780; DB 14; Length 8780;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 8780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 156 CCGGTGAGTACACCGGAATTGCCAGACGACCGGGTCTTTTGTGGATCAACCCGCTCAA 215
 Db 1 CCGGTGAGTACACCGGAATTGCCAGACGACCGGGTCTTTTGTGGATCAACCCGCTCAA 60
 QY 216 TGCCTGGAGATTGGCGTGCCTCCCGCGAGACTGCTAGCCGAGTAGTGTGGTCCGAA 275
 Db 61 TGCCTGGAGATTGGCGTGCCTCCCGCGAGACTGCTAGCCGAGTAGTGTGGTCCGAA 120
 QY 276 AGCCTTGTGGTACTGCTGATAGGGTGTCTGGAGTGCCTCCGGAGGTCTCGTAGACCG 335
 Db 121 AGCCTTGTGGTACTGCTGATAGGGTGTCTGGAGTGCCTCCGGAGGTCTCGTAGACCG 180
 QY 336 TGCACCATGAGCAGCAATCTCTAAACCTCAAGAAAAAACAACCAACGTAACCAACCGCGC 395
 Db 181 TGCACCATGAGCAGCAATCTCTAAACCTCAAGAAAAAACAACCAACGTAACCAACCGCGC 240
 QY 396 CCACAGACGCTCAAGTTCCCGCGCGGTGTGCATCGTTGGTGGAGTTTACCTGTTGCCG 455
 Db 241 CCACAGACGCTCAAGTTCCCGCGCGGTGTGCATCGTTGGTGGAGTTTACCTGTTGCCG 300
 QY 456 CGCGGGGCCCGAGGTGGGTGTGGCGGCGACATAGGAGGCTTCCGAGCGCTCGCAACCT 515
 Db 301 CGCAGGGGCCCGAGGTGGGTGTGGCGGCGACATAGGAGGCTTCCGAGCGCTCGCAACCT 360
 QY 516 CGTGCAGAGGCGACAACTATCCCAAGAGCTCGCGCACCCGAGGCGAGGGCTGGGCTCAG 575
 Db 361 CGTGCAGAGGCGACAACTATCCCAAGAGCTCGCGCACCCGAGGCGAGGGCTGGGCTCAG 420
 QY 576 CCGGGGTACCTTGGGCCCTCTATGGCAATGAGGGCTGGGGTGGGAGGATGGCTCTCG 635
 Db 421 CCGGGGTACCTTGGGCCCTCTATGGCAATGAGGGCTGGGGTGGGAGGATGGCTCTCG 480
 QY 636 TCACCCCGGGCTCCCGGCTAGTTGGGGCCCAACGACCCCGGCTAGCTCGCGTAAC 695
 Db 481 TCACCCCGGGCTCCCGGCTAGTTGGGGCCCAACGACCCCGGCTAGCTCGCGTAAC 540
 QY 696 TTGGGTAAAGTCAATCGATACCTTACATGGGCTTCGCGATCTCATGGGGTACATTCG 755
 Db 541 TTGGGTAAAGTCAATCGATACCTTACATGGGCTTCGCGATCTCATGGGGTACATTCG 600
 QY 756 CTGTCGCGCGCCCCCTAGGGGGCGCTGCGAGGGCTTGGGACGAGGTGCGGGTCTG 815
 Db 601 CTGTCGCGCGCCCCCTAGGGGGCGCTGCGAGGGCTTGGGACGAGGTGCGGGTCTG 660
 QY 816 GAGGACGCGGTAATGCAACAGGAACTTGGCCGGTCTCTCTTCTCTATCTTCTC 875
 Db 661 GAGGACGCGGTAATGCAACAGGAACTTGGCCGGTCTCTCTTCTCTATCTTCTC 720
 QY 876 TTGGCTTGTCTGCTGCTGTTTGCACCACTCCAGCTTCCGCTTATGAAGTCCGCAACGTGTCC 935
 Db 721 TTGGCTTGTCTGCTGCTGTTTGCACCACTCCAGCTTCCGCTTATGAAGTCCGCAACGTGTCC 780
 QY 936 GGGATATACCATGTGCAGAACGACGTGCTCCAACTCAAGCATGTGTATGAGCGCGGAC 995

QY 3156 ATGAAGCTGGGCGCGCTGACAGTACGTAGCTTTATATACCATCTTACCCACTGCGGGAC 3215
 Db ATGAAGCTGGGCGCGCTGACAGTACGTAGCTTTATATACCATCTTACCCACTGCGGGAC 3060
 QY 3216 TGGGGCCACGCGGCGCTTACGAGACCTTGGCGTGGCGTAGAGCCGCTGCTTCTCCGCC 3275
 Db TGGGGCCACGCGGCGCTTACGAGACCTTGGCGTGGCGTAGAGCCGCTGCTTCTCCGCC 3120
 QY 3276 ATGGAGACCAAGTGCATCACTCGGGAGCAGACACCGCTCGTGTGGGACATCATCTTG 3335
 Db ATGGAGACCAAGTGCATCACTCGGGAGCAGACACCGCTCGTGTGGGACATCATCTTG 3180
 QY 3336 GGTCTACCGCTTCGCGCCGACGAGGAGAGAGATATTTTGGACCGCGTGTAGTCTC 3395
 Db GGTCTACCGCTTCGCGCCGACGAGGAGAGAGATATTTTGGACCGCGTGTAGTCTC 3240
 QY 3396 GAAGGGCAAGGGTGGGACCTTTCGCGCCATCAACCGCTTACTCCCAACAAACGCGGGG 3455
 Db GAAGGGCAAGGGTGGGACCTTTCGCGCCATCAACCGCTTACTCCCAACAAACGCGGGG 3300
 QY 3456 GTACTTGGTGTGATCATCACTAGCCTCAGAGCGCGGACAGAACAGGTTCGAAGGGGAG 3515
 Db GTACTTGGTGTGATCATCACTAGCCTCAGAGCGCGGACAGAACAGGTTCGAAGGGGAG 3360
 QY 3516 GTTCAAGTGGTTCCTACCGCAACACATCTTCTCGCGACCTGCATCAACGGCGTGTG 3575
 Db GTTCAAGTGGTTCCTACCGCAACACATCTTCTCGCGACCTGCATCAACGGCGTGTG 3420
 QY 3576 TGGACTGTCTACCATGCGCTGGCTCGAAGACCCCTAGCCGCTCCAAAGTCCCAATCACC 3635
 Db TGGACTGTCTACCATGCGCTGGCTCGAAGACCCCTAGCCGCTCCAAAGTCCCAATCACC 3480
 QY 3636 CAATGTACACCAATGTAGACCTGGACCTGTGGCTGGCAGGCGCCCCCGGGGCGCG 3695
 Db CAATGTACACCAATGTAGACCTGGACCTGTGGCTGGCAGGCGCCCCCGGGGCGCG 3540
 QY 3696 TCCATGACACCATGACGTGTGGCAGCTCGGACCTTTACTTGTTCAGAGACATCTGTAT 3755
 Db TCCATGACACCATGACGTGTGGCAGCTCGGACCTTTACTTGTTCAGAGACATCTGTAT 3600
 QY 3756 GTCAATCCGGTGGCGCGGAGGCGACAGCAGGGGAACTTACTCTCCCGCAGGCCCGTC 3815
 Db GTCAATCCGGTGGCGCGGAGGCGACAGCAGGGGAACTTACTCTCCCGCAGGCCCGTC 3660
 QY 3816 TCTACCTGAAGGCTCTCGGTGTGTCATTCGCTTGGCTTGGCGGACGTCGTGGG 3875
 Db TCTACCTGAAGGCTCTCGGTGTGTCATTCGCTTGGCTTGGCGGACGTCGTGGG 3720
 QY 3876 GTCTTCGGGCTGCTGTGTGACCCGGGGTCCGAGGGGTGGACTTCATACCCGTT 3935
 Db GTCTTCGGGCTGCTGTGTGACCCGGGGTCCGAGGGGTGGACTTCATACCCGTT 3780
 QY 3936 GAGTCTATGAAATACCATCGGTCTCGGTCTTCACAGCAACTCAACCCCGCGCT 3995
 Db GAGTCTATGAAATACCATCGGTCTCGGTCTTCACAGCAACTCAACCCCGCGCT 3840
 QY 3996 GTACCGCAGACATTCGAAGTGGACATCTGCACGCTTACTGCGAGGGCAGAGCAAC 4055
 Db GTACCGCAGACATTCGAAGTGGACATCTGCACGCTTACTGCGAGGGCAGAGCAAC 3900
 QY 4056 AAAGTGGCGGTGCTGTGTGACCCGACCAAGGTACAGGTGCTGCTGAAACCGCTCGGTT 4115
 Db AAAGTGGCGGTGCTGTGTGACCCGACCAAGGTACAGGTGCTGCTGAAACCGCTCGGTT 3960
 QY 4116 GCGGCGACCTTAGGGTTGGGGGTATATGTCCAAAGCAGACCGGTATGCAACCTTAACATC 4175
 Db GCGGCGACCTTAGGGTTGGGGGTATATGTCCAAAGCAGACCGGTATGCAACCTTAACATC 3961
 QY 4176 AGAAGTGGGTAAAGCATTACACCGGGGCTCCATAGGTACTCCACCTATGGCAAG 4235
 Db AGAAGTGGGTAAAGCATTACACCGGGGCTCCATAGGTACTCCACCTATGGCAAG 4080

QY 4236 TTCTTTCGCGAGCGGTGCTTCTGGGGCGGCTATGACATCATATATGTGATGAGTGC 4295
 Db TTCTTTCGCGAGCGGTGCTTCTGGGGCGGCTATGACATCATATATGTGATGAGTGC 4140
 QY 4296 CACTCAACTGACTCGACTACCACTCTTGGGCAATCGGACAGTCTCTGGACCAAGCGAGACG 4355
 Db CACTCAACTGACTCGACTACCACTCTTGGGCAATCGGCACTCTGACCAAGCGAGACG 4200
 QY 4356 GCTGGAGCGGCTCGTGTGCTGCGCACCGGTACACCTCGGGATGGTTACCGTGCCA 4415
 Db GCTGGAGCGGCTCGTGTGCTGCGCACCGGTACACCTCGGGATGGTTACCGTGCCA 4260
 QY 4416 CACCCCAATATCGAGGAATAGCCTGTCCAAATATGGAGATGCCCTTCTATGSCAAA 4475
 Db CACCCCAATATCGAGGAATAGCCTGTCCAAATATGGAGATGCCCTTCTATGSCAAA 4320
 QY 4476 GGCATCCCATTTAGGCCATCAAGGGGGGAGGCTCTCATTTTCTGCCATTTCCAAAGAG 4535
 Db GGCATCCCATTTAGGCCATCAAGGGGGGAGGCTCTCATTTTCTGCCATTTCCAAAGAG 4380
 QY 4536 AATGTGACGAGCTCGCGCAAGCTGACAGGCTCGGACTGAAACCTGTAGCATATATAC 4595
 Db AATGTGACGAGCTCGCGCAAGCTGACAGGCTCGGACTGAAACCTGTAGCATATATAC 4440
 QY 4596 CGGGGCTTGTATGTGCTGCTATACCGCTATCGGAGAGCTGCTTCTGCTGGCAACAGAC 4655
 Db CGGGGCTTGTATGTGCTGCTATACCGCTATCGGAGAGCTGCTTCTGCTGGCAACAGAC 4500
 QY 4656 GCTCTAATGACGGTTCACCGGCGATTTGACTGACTGANTGACTGTGAAATCATGTGTC 4715
 Db GCTCTAATGACGGTTCACCGGCGATTTGACTGACTGANTGACTGTGAAATCATGTGTC 4560
 QY 4716 ACCACAGAGTGCAGCTTGCAGTTGGATCCACCTTCCATTGACAGACGACCGTGGCCC 4775
 Db ACCACAGAGTGCAGCTTGCAGTTGGATCCACCTTCCATTGACAGACGACCGTGGCCC 4620
 QY 4776 CAAGAAGCGGTGTGCGGCTCGCAACCGCGAGGTAGAACTGGCAGGGGTAGGAGTGGCATC 4835
 Db CAAGAAGCGGTGTGCGGCTCGCAACCGCGAGGTAGAACTGGCAGGGGTAGGAGTGGCATC 4680
 QY 4836 TACAGTTGTGACTCCAGAGAAACCGCGCTGGGCAATGTTCTGATCTTCGGTCTCTGTGT 4895
 Db TACAGTTGTGACTCCAGAGAAACCGCGCTGGGCAATGTTCTGATCTTCGGTCTCTGTGT 4740
 QY 4896 GAGTGTATGACGCGGCTGTGCTTGGTATGACTCACCGCGCTGAGACCTCGGTAGG 4955
 Db GAGTGTATGACGCGGCTGTGCTTGGTATGACTCACCGCGCTGAGACCTCGGTAGG 4800
 QY 4956 TTGCGGCTTACCTAAATACACAGGGTTGCGGCTGCGCAGGACCATCTGGAGTTCTGG 5015
 Db TTGCGGCTTACCTAAATACACAGGGTTGCGGCTGCGCAGGACCATCTGGAGTTCTGG 4860
 QY 5016 GAGAGGCTTTCACAGGCTCACCAATAGATGCGCACTTCTGTCCAGACATAACAG 5075
 Db GAGAGGCTTTCACAGGCTCACCAATAGATGCGCACTTCTGTCCAGACATAACAG 4920
 QY 5076 GCAGGAGACAACTTCTCTTACCTGTGGCATATCAAGCTACAGTGTGCGCGGCGTCAA 5135
 Db GCAGGAGACAACTTCTCTTACCTGTGGCATATCAAGCTACAGTGTGCGCGGCGTCAA 4980
 QY 5136 GCTCCACCTCCATCGTGGGACCAAAATGTGGAAGTGTCTCATACGGGTGAAACCTACATG 5195
 Db GCTCCACCTCCATCGTGGGACCAAAATGTGGAAGTGTCTCATACGGGTGAAACCTACATG 5040
 QY 5196 CAGGGGCAACCCCTGTGTATAGGCTAGGACCGTCCAAATGAGTTCATCTCTCACA 5255
 Db CAGGGGCAACCCCTGTGTATAGGCTAGGACCGTCCAAATGAGTTCATCTCTCACA 5100
 QY 5256 CACCCCATACATAATACATCATGGCATGCTCGGCTCACTGAGGTGCTCACTAGC 5315
 Db CACCCCATACATAATACATCATGGCATGCTCGGCTCACTGAGGTGCTCACTAGC 5160
 QY 5316 ACCTGGGTGCTGTAGGCGGAGTCTTTGCAAGCTTTTGGCGCATACTGCTGACGACGAGC 5375

PALSTGLIHLHONIVDVQVLYGVGSFVFAIKWEYILLFLLLADARVCACIMMMLL
 IAQAEALJENLVNLAASVAGAHGILSLVFFPCAAMWIKRGLAPGAAFYGVWPPLL
 LLLALPPRAYALDREMAASCGGNAVIGLVELILSPYKYVFLFELIWLWLYRFLTRAH
 MQWVPLPVRGRDAILILITCAVHPELIFDIKLLALLGLPMLVQAGITVUPFVR
 AQGLIRACMLVREGAGHYIQMAFMKLGALTGYVYNHLLTPURDWAHAGLRDLAVAE
 PVFSANETKVIITWAGDAAACGDIILGLPVSARAKEIFLGPADSLGGGWRLLAPIT
 AVSQTRGVGLVGTITLDRGNQKQVEGVVSTATQCSFLATCINGVCMVTHVHGASK
 TLAGPKGPITOMVTNVDLVLGVCAPEGASMTPCSGSDLVLTVRHADVIPVRGR
 DSGSILSPVUSLVSGSGPLLCPSHGVGVFRAAVCTRGVAKAVLIPVESMETT
 MRSFVTFIDNSPRAVPTQFOVHLHAPTSGSKTKYKPAAYAAQGYKXIVLIPNSVAATL
 GFAYMSKAGIDPNTKGTGTLITGSIYSITKPLADGGCGGAYDIIICDECHS
 TDSITLIGTGLDQDETAGARLWLATATPGSVTPVPHNIBEIGLSNNGEIPFYK
 APTIEAIKGRHILIFCHSKKCDLAAKLTGLGLNAVAYYRGLDVSVIPIGVVVVA
 TDALMTGFTGDSDNCNTVOTQDFSLDPTFTTETTPODAVSRORRGTGRG
 RSGIYRFTPEGRPSGWFDSVLCBVDQCAWELTPAETSVLRLAYLNTPLGVCC
 DHLEFVESFTGLTHIDAFHSQKQAGNFPVLVAYQATVCARAQAPPSMDQWKC
 LIRKPLTHGFTPLRDLGAVQNEVILTHPIIKYIMACMSADLEVVISIWLVGVLVLA
 ALAAVCLTTTSVVIVIRILISLGRPAVVEDREVILYQFDEMEECASOLPYIEQMGLAE
 QYQKALGLLQATKQAEAAAPVESKWALETFAKHMWNFISGQYLAGLSTLFGN
 PATASIMFTASITSPITTOHTLLFNLGCVAAQLAPPASAFVAGAGIAGAAGVSI
 GLKVLVDILAGYGVAGALVAFKVMGVRPSTEDLVNLLPAILSPGALVGVVCAA
 ILRHHVQCEGAVQWNRILAFASRGNHVSPTHYVPSDAAABVTQILSLATITQLK
 RLKQWLNEDCSTPCSGWLRDWDNICTVLIDFKTLQSKLLRPLRGVPLSCQRYK
 GWRGDIIMGTCPCQAQIAGHVKNMSRILVGRPTCSNTWHGTFPINAITYGTPTSP
 APNYSRALWAAEYEVETRVGDFHYVTMTDNDKPCQCPAPPEFTVEDGVRLLR
 YAPACKPLREDFVQVGLNQLVGSOLPCPEPDPVTLTSMLTDPESHITAETAKRL
 ARGPSPLAASSASAPSLKATCTTHDPSDADLIKANLWRQMGNIITRVSBN
 KVVILDSFRLPHAGPERETSVAAEILKRSKPSALPTWARPDYNPPLLESKDPDY
 VPPVHGCPPLPTKAPPIPPRKRPTVLTESNVSSALAEATKTTPSGSSSVDSGT
 ATAPLDIADSDGGDVSIESMPLGEPDPLSDSGWSVTSSEASEDVVCCSMS
 YTWGALITPCAAEESKLPINLSLLRHHNMYATTSRSASLRQKVTFLDQLVD
 DHRDLVKEKAKATVKALLSIEACKLTTPHSKSKFYGAKDVRNLSRAVHL
 RSVNEDLLETETPIDIITMAKGEVFCVQPEKGRKPARLIVPDLGVRVCEKWLID
 VSTLPQAVNGSVGYOFSKORVEPLVNTWKSKCPMPSYDTRCFDPTVTSEDIRV
 ESILYQCCDLAPARQAIRLSIRLYIGGPLTNSKQNCQYRCRCSAGVLTSCGNL
 TCLKAPARAKLQDQTMVNGDDLWICESAGTQEDAAALRAATEAMRVSAPPG
 DPPEYEDLELITCSNSVNSVHADGKRVYILTRDPTTPLARAAETARHTPINSWL
 GNTIMVAPTILWARMILMTHFTFSILLAQEQLKALDC"

ORIGIN

Query Match 90.98; Score 8724; DB 14; Length 8780;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 8745; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 156 CCGGTGAGTACACCGGAATTGCCAGGACGACCGGGTCTCTTTGGATCAACCGCTCAA 215
 Db 1 CCGGTGAGTACACCGGAATTGCCAGGACGACCGGGTCTCTTTGGATCAACCGCTCAA 60

QY 216 TGCCTGGAGATTGGCGTCCCCCGCGAGACTCTAGCCAGTAGTGTGGGTGCGGAA 275
 Db 61 TGCCTGGAGATTGGCGTCCCCCGCGAGACTCTAGCCAGTAGTGTGGGTGCGGAA 120

QY 276 AGGCTTGTGCTACTGCTGATAGGTGCTTGGAGTGCCCGGAGGTCTCGTAGACCG 335
 Db 121 AGGCTTGTGCTACTGCTGATAGGTGCTTGGAGTGCCCGGAGGTCTCGTAGACCG 180

QY 336 TGCACCATGAGCAGATCTCTAAACCTCAAGAAACCAACGTAACACCAACCGCGC 395
 Db 181 TGCACCATGAGCAGATCTCTAAACCTCAAGAAACCAACGTAACACCAACCGCGC 240

QY 396 CCACAGACGTCAGATTGCCGGGCGGTGCTGATGCTGCTAGCGAGTAGTGTGGGTGCGG 455
 Db 241 CCAAGGACGTCAGATTGCCGGGCGGTGCTGATGCTGCTAGCGAGTAGTGTGGGTGCGG 300

QY 456 CGCAGGCGCCCGAGGTGGGTGTCGGCGACTAGGAGGCTCCGAGCGGTGCGCAACCT 515
 Db 301 CGCAGGCGCCCGAGGTGGGTGTCGGCGACTAGGAGGCTCCGAGCGGTGCGCAACCT 360

QY 516 CGTGEAAGGCGACAACTATCCCAAAGGCTCGCCGACCCGAGGCGAGGCTGGGCTCAG 575
 Db 361 CGTGEAAGGCGACAACTATCCCAAAGGCTCGCCGACCCGAGGCGAGGCTGGGCTCAG 420

QY 576 CCGGGGTACCTTGGCCCTCTATGGGCAATGAGGCGCTGGGGTGGGCGAGATGCTCCTG 635

Db 421 CCGGGGTACCTTGGCCCTCTATGGCAATGAGGCGCTGGGGTGGCGAGATGGCTCCTG 480
 QY 636 TCACCCCGGGCTCCGGGCTAGTTGGGGCCCCACGAGCCCCCGGGTAGGTGCGGTAA 695
 Db 481 TCACCCCGGGCTCCGGGCTAGTTGGGGCCCCACGAGCCCCCGGGTAGGTGCGGTAA 540
 QY 696 TTGGGTAAAGGTCAATGATACCTTATACATGCGGCTTTCGCGATCTCATGGGGTACATTCG 755
 Db 541 TTGGGTAAAGTCAATGATACCTTATACATGCGGCTTTCGCGATCTCATGGGGTACATTCG 600

QY 756 CTCGTGCGGCGCCCTAGGGGGCGGTGCGAGGGCTTGGGACACGTTGTCGGGTCTG 815
 Db 601 CTCGTGCGGCGCCCTAGGGGGCGGTGCGAGGGCTTGGGACACGTTGTCGGGTCTG 660

QY 816 GAGGAGGGCGTGAATATATGCAACAGGAACTTGCCTGGGTGCTCTTCTATCTCTC 875
 Db 661 GAGGAGGGCGTGAATATATGCAACAGGAACTTGCCTGGGTGCTCTTCTATCTCTC 720

QY 876 TTGGCTCTCTCTCTCTCTTGAACCTCCAGCTTCCGCTTATGAAAGTGGCAACGTGTC 935
 Db 721 TTGGCCCTCTCTCTCTTGTGACCATCCAGCTTCCGCTTATGAAAGTGGCAACGTGTC 780

QY 936 GGGATATACCTGTGTCAGAACGACTGCTTCAACTCAAGCAITGTGTATGAGGCGAGGAC 995
 Db 781 GGGATATACCTGTGTCAGAACGACTGCTTCAACTCAAGCAITGTGTATGAGGCGAGGAC 840

QY 996 GTGATCATCATCTCCCGGGTGGGTGCTGCTGTTCAGGAGGTAACAAGCTCCCGTTCG 1055
 Db 841 GTGATCATCATCTCCCGGGTGGGTGCTGCTGTTCAGGAGGTAACAAGCTCCCGTTCG 900

QY 1056 TGGGTAGCGCTCACTCCCACTGCGCGGCGAGGAATGCGAGCGTCCCACTACGACAATA 1115
 Db 901 TGGGTAGCGCTCACTCCCACTGCGCGGCGAGGAATGCGAGCGTCCCACTACGACAATA 960

QY 1116 CGACGCCAGCTGACTTGTCTGTTGGGAGCGGTGCTTCTGCTCCGCTATGATAGTGGG 1175
 Db 961 CGACGCCAGCTGACTTGTCTGTTGGGAGCGGTGCTTCTGCTCCGCTATGATAGTGGG 1020

QY 1176 GATCTCTCGGATCTATATTTCTCTGCTCCAGCTGTTTCACTTCTGCGCTCCCGGCAAT 1235
 Db 1021 GATCTCTCGGATCTATATTTCTCTGCTCCAGCTGTTTCACTTCTGCGCTCCCGGCAAT 1080

QY 1236 GAGACAGTCAGGACTGCAACTGCTCAATCTATCCCGGCAATGATACGTCACGCGATG 1295
 Db 1081 GAGACAGTCAGGACTGCAACTGCTCAATCTATCCCGGCAATGATACGTCACGCGATG 1140

QY 1296 GCTTGGGATATGATGATGAACTGCTCACTCAACAGCCCTAGTGGTTCGAGTTGCTC 1355
 Db 1141 GCTTGGGATATGATGATGAACTGCTCACTCAACAGCCCTAGTGGTTCGAGTTGCTC 1200

QY 1356 CGGATCCCAACAGCTGCTGCTGGACATGTTGGCGGGGGCCACTGGGGAGTCTTGGCGG 1415
 Db 1201 CGGATCCCAACAGCTGCTGCTGGACATGTTGGCGGGGGCCACTGGGGAGTCTTGGCGG 1260

QY 1416 CTTGCTCTATATTTCCATGAGTGGGAACTGGGTGATAGTTCTGATTTGTGGCGTACTCTTT 1475
 Db 1261 CTTGCTCTATATTTCCATGAGTGGGAACTGGGTGATAGTTCTGATTTGTGGCGTACTCTTT 1320

QY 1476 GCGCGGTTTCAGGGGAGACCCACAGAGGGGAGGTGCGCGCCACACCTCCGG 1535
 Db 1321 GCGCGGTTTCAGGGGAGACCCACAGAGGGGAGGTGCGCGCCACACCTCCGG 1380

QY 1536 TTCACTCTCTCTTCTCATCTGGGGCTTCAGAAAAATCCAGTTGTGAATACCAACGCG 1595
 Db 1381 TTCACTCTCTTCTCATCTGGGGCTTCAGAAAAATCCAGTTGTGAATACCAACGCG 1440

QY 1596 AGTGTGCATCAACAGGACTGCTTAAATGCAATGATGCTTCAATCCAACTGGGTTCTTT 1655
 Db 1441 AGTGTGCATCAACAGGACTGCTTAAATGCAATGATGCTTCAATCCAACTGGGTTCTTT 1500

QY 1656 GCGCGCTGTTTACGACACAGTTCACTGCTCGGGTGGCCCGGAGCGATGCGCAGC 1715

Db 5881 GCAGCAATCTGCGTCGGCACTGGGCCC CGGAGAGGGGGCTGTGCACTGGATGAACCGG 5940
Qy
Qy 5096 CTCATAGCGTTCCTTCGCGGGTAACCACTGTCCTCCCTACGACACTATGTGCTGAGAGC 6155
Db 5941 CTGATAGCGTTCCTTCGCGGGTAACCACTGTCCTCCCTACGACACTATGTGCTGAGAGC 6000
Qy 6156 GAGCGTGCAGACAGTGTCACTCAGATCCTCTCTAGCCTTACCATCACTCAACTGCTGAAG 6215
Db 5001 GAGCGTGCAGACAGTGTCACTCAGATCCTCTCTAGCCTTACCATCACTCAACTGCTGAAG 6060
Qy 5216 CGGCTCCACAGTGGATTAAGAGGATGTCTCTACGCCATCTCTCCGCTCGTGGCTAAGG 6275
Db 5061 CGGCTTCACCACTGGATTAAGAGGATGTCTCTACGCCATCTCTCCGCTCGTGGCTAAGG 6120
Qy 6276 GATGTTGGGATTTGGATATCAGAGTGTGACTGACTTCAAGACTTGGCTCCAGTCCAA 6335
Db 5121 GATGTTGGGATTTGGATATCAGAGTGTGACTGACTTCAAGACTTGGCTCCAGTCCAA 6180
Qy 6336 CTCTCTCGCGGTTACCGGGAGTCCCTTTCTCTGTATGCCAAACCGGGGTACAAGGAGTC 6395
Db 6181 CTCTCTCGCGGTTACCGGGAGTCCCTTTCTCTGTATGCCAAACCGGGGTACAAGGAGTC 6240
Qy 6396 TGGCGGGGAGCGGATCATGCAACCACTGTCCTGACGCTGCGAGACAGATGCGCGGACAT 6455
Db 6241 TGGCGGGGAGCGGATCATGCAACCACTGTCCTGACGCTGCGAGACAGATGCGCGGACAT 6300
Qy 6456 GTCAAAAACGGTTCCATAGAGATCGTAGGCGCTTAGACCTGCAACACAGTGGCGGCA 6515
Db 6301 GTCAAAAACGGTTCCATAGAGATCGTAGGCGCTTAGACCTGCAACACAGTGGCGGCA 6360
Qy 6516 ACGTTCCCACTAACCGATACACACAGGACCTTTGCAACCCCTCCCGGGGCCCAACTAT 6575
Db 6361 ACGTTCCCACTAACCGATACACACAGGACCTTTGCAACCCCTCCCGGGGCCCAACTAT 6420
Qy 6576 TCACAGGCGCTATGGCGGTGGCTGCTGAGAGTACGTGAGGTACGTGGGTGGGGAT 6635
Db 6421 TCACAGGCGCTATGGCGGTGGCTGCTGAGAGTACGTGAGGTACGTGGGTGGGGAT 6480
Qy 6636 TTCACCTACGTGACGGCATGACACATGCAACCTAAAGTGCCATGCCAGGTTCGGGCC 6695
Db 6481 TTCACCTACGTGACGGCATGACACATGCAACCTAAAGTGCCATGCCAGGTTCGGGCC 6540
Qy 6696 CCCGAATTTCTACCGAGGTGGATGGAGTGGCTGCAACAGTACGCTCCGGGTGCAAA 6755
Db 6541 CCCGAATTTCTACCGAGGTGGATGGAGTGGCTGCAACAGTACGCTCCGGGTGCAAA 6600
Qy 6756 CCTCTCTACGGAGGACGTCACTTCAGGTGGGCTCAACATACCTTGGTGGGTGCG 6815
Db 6601 CCTCTCTACGGAGGACGTCACTTCAGGTGGGCTCAACATACCTTGGTGGGTGCG 6660
Qy 6816 CAGCTCCATGCGAGCCGAAACCGAGCTAACAGTGTCTTCTCCATGCTCACCGATCCC 6875
Db 6661 CAGCTCCATGCGAGCCGAAACCGAGCTAACAGTGTCTTCTCCATGCTCACCGATCCC 6720
Qy 6876 TCCACATTAACAGAGAGACGCTTAACGCTAGGCTGCTGAGGCTTCCCCCTCTTTA 6935
Db 6721 TCCACATTAACAGAGAGACGCTTAACGCTAGGCTGCTGAGGCTTCCCCCTCTTTA 6780
Qy 6936 GCCAGCTCATAGCTAGCAGTGTCTCTGGCTTCTTTGAGGCGCATGCACTACCCAC 6995
Db 6781 GCCAGCTCATAGCTAGCAGTGTCTCTGGCTTCTTTGAGGCGCATGCACTACCCAC 6840
Qy 6996 CATGACTCCCCGAGCGCTCATCATGAGGCCAACCTCTGTGGCGGCGAGGATGGGC 7055
Db 6841 CATGACTCCCCGAGCGCTCATCATGAGGCCAACCTCTGTGGCGGCGAGGATGGGC 6900
Qy 7056 GGAAACATCACTCCGTTGGAGTACAGAAATAGTAAATTTCTGACTCTTTTCCGAACCG 7115
Db 6901 GGAAACATCACTCCGTTGGAGTACAGAAATAGTAAATTTCTGACTCTTTTCCGAACCG 6960
Qy 7116 CTTTCCCGGAGGGGATGAGAGGAGATATCCCTCCGGCGGAGATCTTCCGAAATCC 7175
Db 6961 CTTTCCCGGAGGGGATGAGAGGAGATATCCCTCCGGCGGAGATCTTCCGAAATCC 7020

Qy 7176 AGGAAGTTCCCTCAGCGTTGCCATATGGSCACGCCCGACTACATCTCCACTGCTA 7235
Db 7021 AGGAAGTTCCCTCAGCGTTGCCATATGGSCACGCCCGACTACATCTCCACTGCTA 7080
Qy 7236 GAGTCTCGAAGACCCGACTACGTCCTCCGCTGGTACACGGATGCCATTTGCCACT 7295
Db 7081 GAGTCTCGAAGACCCGACTACGTCCTCCGCTGGTACACGGATGCCATTTGCCACT 7140
Qy 7296 ACCAAGCTCTTCCAAATACCACTCCACGGAAGAGAGACGGTTGTCTGTACAGATCC 7355
Db 7141 ACCAAGCTCTTCCAAATACCACTCCACGGAAGAGAGACGGTTGTCTGTACAGATCC 7200
Qy 7356 AATGTGCTTCTGCTTGGCGAGCTCGCACTTAAGACCTTGGTAGCTCCGATCGTCG 7415
Db 7201 AATGTGCTTCTGCTTGGCGAGCTCGCACTTAAGACCTTGGTAGCTCCGATCGTCG 7260
Qy 7416 GCCGTTGATAGCGGACCGGACCGGCTTCTGTGACCTGGCTCCGACGAGGTGACAA 7475
Db 7261 GCCGTTGATAGCGGACCGGACCGGCTTCTGTGACCTGGCTCCGACGAGGTGACAA 7320
Qy 7476 GGAATCCGAGCTTGAGTCTGACTCTTCCATGCGCCCTTGAAGGGAGCGGGGACCC 7535
Db 7321 GGGTCCGAGCTTGAGTCTGACTCTTCCATGCGCCCTTGAAGGGAGCGGGGACCC 7380
Qy 7536 GATCTCAGCGACGGGTCTTGGTCTACCGTGAAGTGAAGGCTAGTGAGGATGTCTGTCG 7595
Db 7381 GATCTCAGCGACGGGTCTTGGTCTACCGTGAAGTGAAGGCTAGTGAGGATGTCTGTCG 7440
Qy 7596 TGTCTAATGTCTTATACGTGGACGAGCGGCTGTATCACGCCATGCGCTCGGAGGAAAT 7655
Db 7441 TGTCTAATGTCTTATACGTGGACGAGCGGCTGTATCACGCCATGCGCTCGGAGGAAAT 7500
Qy 7656 AAGCTCCCATCAACCGGTTGAGACTCTTGTCTGGCTCACACAACTGCTCTAGCC 7715
Db 7501 AAGCTCCCATCAACCGGTTGAGACTCTTGTCTGGCTCACACAACTGCTCTAGCC 7560
Qy 7716 ACAACATCCCGACGCAAGCCCTCCGCGAAGAAAGTCACTTTGACAGATTTGCAAGTC 7775
Db 7561 ACAACATCCCGACGCAAGCCCTCCGCGAAGAAAGTCACTTTGACAGATTTGCAAGTC 7620
Qy 7776 CTGATGATCATTAACCGGACGTTACTCAAGGAGATGAAGCGGCTCCACAGTTAAG 7835
Db 7621 CTGATGATCATTAACCGGACGTTACTCAAGGAGATGAAGCGGCTCCACAGTTAAG 7680
Qy 7836 GCTAAGCTTCTATATAGAGGAGCTGCAAGCTGACGCCGCCACATTTGGGCCAATCC 7895
Db 7681 GCTAAGCTTCTATATAGAGGAGCTGCAAGCTGACGCCGCCACATTTGGGCCAATCC 7740
Qy 7896 AAATTTGGCTATGGGCAAGGACGTCGGAAACCTATCCAGCAGGGCCGTTAACCAATC 7955
Db 7741 AAATTTGGCTATGGGCAAGGACGTCGGAAACCTATCCAGCAGGGCCGTTAACCAATC 7800
Qy 7956 CGTCCGTTGGGAGGACCTGCTGGAAGACACTGAAACCACTTGAACCACTGATG 8015
Db 7801 CGTCCGTTGGGAGGACCTGCTGGAAGACACTGAAACCACTTGAACCACTGATG 7860
Qy 8016 GCAAAAGTGAAGTTTCTGCTCCAGCAGAGAGGAGGCGGCGAGCGCTCCCTT 8075
Db 7861 GCAAAAGTGAAGTTTCTGCTCCAGCAGAGAGGAGGCGGCGAGCGCTCCCTT 7920
Qy 8076 ATCGTATTTCCAGACCTGGGAGTCTGTGTATGCGAAGATGGCCCTTTTACGACGTGGTC 8135
Db 7921 ATCGTATTTCCAGACCTGGGAGTCTGTGTATGCGAAGATGGCCCTTTTACGACGTGGTC 7980
Qy 8136 TCCACCTCTCTCAGCGCGTGTATGGGCTCTCTATAACGAAATTTCAATCTCCCGGAGAG 8195
Db 7981 TCCACCTCTCTCAGCGCGTGTATGGGCTCTCTATAACGAAATTTCAATCTCCCGGAGAG 8040
Qy 8196 CGGGTGAAGTCTCTGCTGATACCTGGAATCAAGAAATGCGCCTATGGCTTCTCATAT 8255
Db 8041 CGGGTGAAGTCTCTGCTGATACCTGGAATCAAGAAATGCGCCTATGGCTTCTCATAT 8100

```
QY 8256 GACACCCGCTGTTTGAACCTCAACGGTCACTGAGAGTGACATTGCTGTTGAGAGTGCAATT 8315
|
|
|
Db 8101 GACACCCGCTGTTTGAACCTCAACGGTCACTGAGAGTGACATTGCTGTTGAGAGTGCAATT 8160
|
|
|
QY 8316 TACCAATGTTGACCTTGGCCCCGAGGCGCAGACAGGCCATAGATGCTGCTCAGAGAGCGG 8375
|
|
|
Db 8161 TACCAATGTTGACCTTGGCCCCGAGGCGCAGACAGGCCATAGATGCTGCTCAGAGAGCGG 8220
|
|
|
QY 8376 CTTTACATCGGGGGTCCCTGACTCAACTCAAAAGGCGCAGAACTGCGGTATATCCCGGGTGC 8435
|
|
|
Db 8221 CTTTACATCGGGGGTCCCTGACTCAACTCAAAAGGCGCAGAACTGCGGTATATCCCGGGTGC 8280
|
|
|
QY 8436 CGCGCAAGTGGCGTCTGCTGACGACTAGCTGCGGTAATACCCCTCACATGTTACTTGAAGGCC 8495
|
|
|
Db 8281 CGCGCAAGTGGCGTCTGCTGACGACTAGCTGCGGTAATACCCCTCACATGTTACTTGAAGGCC 8340
|
|
|
QY 8496 ACTGACGCTGTCGAGCTGCAAAAGCTCCAGGACTCCAGATGCTCGTGACGAGACGAC 8555
|
|
|
Db 8341 ACTGCGGCTGTCGAGCTGCAAAAGCTCCAGGACTCCAGATGCTCGTGACGAGACGAC 8400
|
|
|
QY 8556 CTTTGTGTTATCTGTGAAGCGCGGGAACCCAGGAGGATGCGGCGGCGCTTACGAGCCTTC 8615
|
|
|
Db 8401 CTTTGTGTTATCTGTGAAGCGCGGGAACCCAGGAGGATGCGGCGGCGCTTACGAGCCTTC 8460
|
|
|
QY 8616 ACGGAGGCTATGACTAGGATTTCCGCCCCCCCCCGGGATCCGCCCCCAACAGAAATACGAC 8675
|
|
|
Db 8461 ACGGAGGCTATGACTAGGATTTCCGCCCCCCCCCGGGATCCGCCCCCAACAGAAATACGAC 8520
|
|
|
QY 8676 CTGGAGCTGATAACATCATGTTCTCCATGTTGTCAGTCCGCGCAGCAGATGCTATCGGCAA 8735
|
|
|
Db 8521 CTGGAGCTGATAACATCATGTTCTCCATGTTGTCAGTCCGCGCAGCAGATGCTATCGGCAA 8580
|
|
|
QY 8736 AGGGATATCTACTCCTACCCGCTGACCCCAACCCCGCTTGGACCGGCTCGGTGGGAGACA 8795
|
|
|
Db 8581 AGGGATATCTACTCCTACCCGCTGACCCCAACCCCGCTTGGACCGGCTCGGTGGGAGACA 8640
|
|
|
QY 8796 GTAGACACACTCCAAATCACTTCTGGTAGGCAATATCATATGATGCGGCCACCCCTA 8855
|
|
|
Db 8641 GCTAGACACACTCCAAATCACTTCTGGTAGGCAATATCATATGATGCGGCCACCCCTA 8700
|
|
|
QY 8856 TGGGCAAGGATCATGCTGATGACTCACTTTTCTCCATCTTCTAGCTCAAGACCACTT 8915
|
|
|
Db 8701 TGGGCAAGGATCATGCTGATGACTCACTTTTCTCCATCTTCTAGCTCAAGACCACTT 8760
|
|
|
QY 8916 GAAAAGCCCTGGATTGTCA 8935
|
|
|
Db 8761 GAAAAGCCCTGGATTGTCA 8780
|
|
|
RESULT 10
LOCUS AF054252 8780 bp RNA linear VRL 02-MAY-1998
DEFINITION Hepatitis C virus HC-J4, pcV-J412 polyprotein gene, partial cds.
ACCESSION AF054252
VERSION AF054252.1 GI:3098641
KEYWORDS Hepatitis C virus
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 8780)
AUTHORS Yanagi,M., St Claire,M., Shapiro,M., Emerson,S.U., Purcell,R.H. and
Bukh,J.
TITLE Transcripts of a chimeric cDNA clone of hepatitis C virus genotype
1b are infectious in vivo
JOURNAL Virology 244 (1), 161-172 (1998)
MEDLINE 98240944
PUBMED 9581788
REFERENCE 2 (bases 1 to 8780)
AUTHORS Yanagi,M. and Bukh,J.
TITLE Direct Substitution
JOURNAL Submitted (18-MAR-1998) Hepatitis Viruses Section, LID, NIAID, NIH,
7 Center Dr., Building 7, Room 201, Bethesda, MD 20892, USA
```

FEATURES
source

```
Location/Qualifiers
1. .8780
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/strain="HC-J4"
/isolate="pcV-J412"
/db_xref="taxon:11103"
187. .->8780
```

CDS

```
/codon_start=1
/product="polyprotein"
/protein_id="AA015726.1"
/db_xref="GI:3098642"
translation="MSTNPKPQKTKNTNRRPDVKFPGGGIUGGVYLLPRGPRL
SRATKASRSQPRGRQIPIKARPEGRMAQPGYPWPLYNEGILWAGVLLSPRG
SVTSWGTDPRRSRNLKVIDITGCFADLMGYIPLVAGAPLGAARALAHGVRLSD
GVNATLPCGCSFISLLALLSLTIPASAYEVNRYSGIYVNDGNSIIVVBAAD
VYHTPCGQVOGSSRCWALPTLAENASVPTTIRREVDLLVGTAAFCASMY
VGLTGGIFLVSQLPTSPRHEHVQDYNCSIFGVSHRMADWMWNSPTALVV
SQLLRIPQAVDMVAGAHGVLAGLAIYSWGNWAKVLIIVALLPAGVDTGHTTGRVA
HTTSGFTSLFSSGAKIQLVNTGNSHWINRTALNCDSLOTGFFAALFYAHKFSH
GTERMAACRPIDWFAQWGPITYTKENSDQRPQWYAPPCGVVPAQVQCGPVYC
FTSPVVGTTRDSGVPYTSWGENETDMLLNTRPPQGNWFGCTWNSSTGFTKCGG
PPNIGVGNETLICTDCEKHPKPEATYTKGSGPMLTPRLVDYPRVLYWHPCTLNF
SIPKRVNMGVGEHRLNAACNWTRGERCNLEDRSLSPLLLSTTEWQILPCAFTLL
PALSTGHLHQNIVDQYLVGVSAFVSFAIKWEYILLLFLLLADARVACALWMLL
IAQEAALENLVLNAAASVAGAHGILSFLVFFCAAYIKGRLAPGAAYFVGPWPLL
LLALPRAYALDREMAASCGAVLGLVFLTSLPYKVFLLTLMWQYFIIRAEAH
MQWVPLNVRGGRDAIILITCAVPELIPDITKLIALILGLPLMLQAGITRVPYFVR
AOGILACMLVRKVAGGHIQMAFMKLGALTGVVYNHLLTPRDWAHAGRLDAVVE
FVPSMETKVTWGAATAAGDILGLPYSAERGKEIFLGPADSLGSGQWRLLAPIT
AYSQTPGVGCVTISLTGRDKNOVEGPEVQVSTATQSFATCLNGVCTVHYGASGK
TLAPGKPIQMTYNTVDLGVQAPPEAKSMPCSCSSDLIVLTHADVIIVRRRG
DSRGLSPSPVSLKSGSGPLLCPSGHVGVVFRAAVCTRGAKAVDYFIPVSMETL
MRVFTDNPSSPAVPQTSVAHLHAPTQKSTKVPAAAYAGYKVLINPVSAAITL
GFGAMSKAGHIDNPITGVRTITGGSITVSTYKGLADGCGGGAAYDIIICDECHS
TDSITLIGITVLDOARTAGARLVLATATPGSVTVPHNIBELIGLSNNGEIPFYCK
AIFLEAKGRHLIFCHSKKCDLAKLTGLGLNNAVAYRGLDVSVIPIGDVVVVA
TDALMTGTFDSDVNTCVTDVFDISLDTFTIETITVPQDAVRSQRRTGKRG
RSGLYFVFTPGERSGSDSVLCCEYDAGCAWYELTFAETSVRLRAYLITGHPVQO
DHLRFWSVFTGLTHIDAFLSQTKAGDNFPLVAYQATVCAQAAPPSPWDMQMK
LIRKPLTGHPTLLYRLGAVONEVILTHPTIKYIMACMDALEVSTWVPGGVLA
ALAACTTGSVTVGRIILSGKPAVVDRELVYQEPDEMEECASQLPYIBQWQLAR
QKQKALGLIQTAKQAEAPVVESSWRALETFWAKMNFISGIQYLAGLSLTGPN
PAIASALFASITSPUTHTLLNLGGWAAQALAPPASAASAFVAGAGAAVGSIT
PAIAGLVILLAGYAGAGALVAFKVMSEVPSTEDVLLLPAILSPALVGVVCAA
ILRRHVGECAVQWMMNLLAFASGRNHVSPTHYVPESDAAARVTLSSLTITQLK
RLHOWINEDSTPCSGSLRDVMDWICTVLTDFKTLQSKLLPLPVPFLSCORGK
GWRSGDGMOTTCGQAQIAGHVKNMIRIVGPRTCNTWHTGTPINAYTIGPCTPSP
APNYSRALRVAAEYEVTRVGDPHYVTGNTDNVCKPCQVPAPETFDVGVRLHR
YAPACKPRLREDVTFQVLAQYLVGSQLPCEPFPDVTITSMLDTHSHITATYAKRL
ARGSPPLASSASQLGAPLAKATCTTHSDPADLIEANLLMQEMGNITRVESEN
KVVIDLFEPLHAGDEREISVAEILRKSKFPSPALFIWAPDYNPFLLESWKDPY
VVPVHGCPPTKAPPTPPPRRKTIVLTESNVSAALAEIATKTFGSGSSVDSGT
ATAPDLASDDGDKSDVSYSSMPLGLEGPDPDLSDGSMSTVSBASDVVCCWS
YTWGALLTPCAABESKILPNEPLSLLRHENMYATISSASLRQKVTDFRQVLDL
DHYRDLKMKAKASITVAKLLSIEACKLTPPHSAKSKFGYGAQVKNLSRAVNH
RSVMEDLLEDTEPIDTIMAKSVFCVQPERGKRKPARLIVFPDLGVKRCENKALIV
VSVLPQAVMGSSYGFQYSPQRVEFLVNTWKSKCPMGFSYDTRFSTVEDIRV
BESIYQCDLAFAPKQAIRSLTERLYIGGPLTNSKQNGCYRRCRASGLTSCNTL
TCYLKATACRAACLODCTMLVNGDLVVICESAGTQBDAAALRAFTEAMTRYAPPG
DPPQPYDELITITSSSNVSAHDASGRVYVLRDPTTPLARAAMETARTPTVNSWL
GNIMYAPTIWARMILMTHFFSILLAQBLEKALDC"
```

ORIGIN

```
Query Match 90.7%; Score 8703.2; DB 14; Length 8780;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 8732; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 156 CCGGTGAGTACACCGGAATTCAGACGACCGGGTCTCTTCTTGGATCAACCCGCTCAA 215
|
|
|
Db 1 CCGGTGAGTACACCGGAATTCAGACGACCGGGTCTCTTCTTGGATCAACCCGCTCAA 60
|
|
|
QY 216 TGCCTGGAGATTGGCGGTGCGCCCGGAGACTGCTAGCCGAGTAGTCTTGGTGGCAA 275
```

Db 61 TGCCTGGAGATTGGGCTGCCCCCGCGAGCTGCTAGCCAGGTAGTGTGGGTCCGGAA 120
 QY 276 AGGCTTGTGCTACTGCTGATAGGTGCTTGCGAGTGCCCGGGAGGTCTCTGTAACCG 335
 Db 121 AGGCTTGTGCTACTGCTGATAGGTGCTTGCGAGTGCCCGGGAGGTCTCTGTAACCG 180
 QY 336 TGCACCATGAGCAGCAATCTTAACCTCAAGAAACCAAAACCAACCAACCAACCGCGC 395
 Db 181 TGCACCATGAGCAGCAATCTTAACCTCAAGAAACCAAAACCAACCAACCAACCGCGC 240
 QY 396 CACAGACCTCAAGTTCCCGGGCGGTGCTGAGTTCAGATCGTTGAGTTTACCTTTCOCG 455
 Db 241 CACAGACCTCAAGTTCCCGGGCGGTGCTGAGTTCAGATCGTTGAGTTTACCTTTCOCG 300
 QY 456 CGCAGGGCCCCAGGTTGGGTGCGCGCAGCTAGGAAGCTTCCGAGCGGTGCGACCT 515
 Db 301 CGCAGGGCCCCAGGTTGGGTGCGCGCAGCTAGGAAGCTTCCGAGCGGTGCGACCT 360
 QY 516 CGTGGAAAGGCGCAACCTATCCCAAGGCTCGCGACCCGAGGCGAGGCTGGGCTCAG 575
 Db 361 CGTGGAAAGGCGCAACCTATCCCAAGGCTCGCGACCCGAGGCGAGGCTGGGCTCAG 420
 QY 576 CCGGGTACCTTGGGCCCTCTATGGCAATGAGGGCTGGGGTGGGAGGATGCTCTG 635
 Db 421 CCGGGTACCTTGGGCCCTCTATGGCAATGAGGGCTGGGGTGGGAGGATGCTCTG 480
 QY 636 TCCGCCCGCGCTCCCGGCTAGTTGGGGCCCGACGACCCCGGCGGTAGGTGCGTAAC 695
 Db 481 TCACCCCGCGCTCCCGGCTAGTTGGGGCCCGACGACCCCGGCGGTAGGTGCGTAAC 540
 QY 696 TTGGTTAAGGTATCATGATACCTTACATGCGGCTTCGCCGATCATGGGTATCATTCOG 755
 Db 541 TTGGTTAAGGTATCATGATACCTTACATGCGGCTTCGCCGATCATGGGTATCATTCOG 600
 QY 756 CTGCTCGGCGCCCTAGGGGGCGCTGCGAGGGCCCTTGGCAACCGGTGTCGGGTCTG 815
 Db 601 CTGCTCGGCGCCCTAGGGGGCGCTGCGAGGGCCCTTGGCAACCGGTGTCGGGTCTG 660
 QY 816 GAGGACGCGTGAATGCAACAGGGAATTCGCCGGTTCCTTCTCTATCTCTCTC 875
 Db 661 GAGGACGCGTGAATGCAACAGGGAATTCGCCGGTTCCTTCTCTATCTCTCTC 720
 QY 876 TTGGCTGTGCTGCTGTTGACCATCCAGCTTCGCTTATGAAGTGGCAAGCTGTC 935
 Db 721 TTGGCTGTGCTGCTGTTGACCATCCAGCTTCGCTTATGAAGTGGCAAGCTGTC 780
 QY 936 GGGATATACCATGCAAGACGCTGCTCAGCTCAAGCAATGCTATGAGGCGAGCGAC 995
 Db 781 GGGATATACCATGCAAGACGCTGCTCAGCTCAAGCAATGCTATGAGGCGAGCGAC 840
 QY 996 GTGATCAFGCATCTCCCGGTGCGTGCCTGTGTTGAGGAGGTAAAGCTCCCGTTC 1055
 Db 841 GTGATCATGATATCCCGGTGCGTGCCTGTGTTGAGGAGGTAAAGCTCCCGTTC 900
 QY 1056 TGGTATGCGCTCACTCCACGCTCGCGGCCAGGAATGCCAGCTCCCGACTACGACAATA 1115
 Db 901 TGGTATGCGCTCACTCCACGCTCGCGGCCAGGAATGCCAGCTCCCGACTACGACAATA 960
 QY 1116 CGAGCCACGCTCGACTGCTGTTGGAGCGGCTGCTTCTGCTCGCTATGTAAGTGGGG 1175
 Db 961 CGAGCCACGCTCGACTGCTGTTGGAGCGGCTGCTTCTGCTCGCTATGTAAGTGGGG 1020
 QY 1176 GATCTCTCGGATCTATTTCTCTGCTCCAGCTGTTCACTTCTCGCTTCGCGGAT 1235
 Db 1021 GATCTCTCGGATCTATTTCTCTGCTCCAGCTGTTCACTTCTCGCTTCGCGGAT 1080
 QY 1236 GAGACATGCGAGGCTCACTGCTCAATCTATCCCGGCCATGATACAGTCAACCGCATG 1295
 Db 1081 GAGACATGCGAGGCTCACTGCTCAATCTATCCCGGCCATGATACAGTCAACCGCATG 1140
 QY 1296 GCTTGGATATGATGATGAACTGCTACCTACACAGCCCTAGTGGTCCAGTTGCTC 1355

Db 1141 GCTTGGGATATGATGATGAACCTGCTCACTCAACAGCCCTAGTGTGTCGAGTTGCTC 1200
 QY 1356 CGGATCCCAAGAGCTCTGCTGAGACATGTTGGGGGGCCACCTGGGAGTCTCTGGCGGGC 1415
 Db 1201 CGGATCCCAAGAGCTCTGCTGAGACATGTTGGGGGGCCACCTGGGAGTCTCTGGCGGGC 1260
 QY 1416 CTTGCTTACTATTCTCCATGTTAGGGAATGCGGTAAAGTTCTGATTTGCGGCTACTCTTT 1475
 Db 1261 CTTGCTTACTATTCTCCATGTTAGGGAATGCGGTAAAGTTCTGATTTGCGGCTACTCTTT 1320
 QY 1476 GCCGGCTTGAACGGGAGAGCCACACAGACGGGAGGTTGGCGGCGCACACCACTCGGG 1535
 Db 1321 GCCGGCTTGAACGGGAGAGCCACACAGACGGGAGGTTGGCGGCGCACACCACTCGGG 1380
 QY 1536 TTCAGTTCCTTTCTCATCTGGGGCTCTCAAAAATCCAGTTCTGAAATCAACAACGGC 1595
 Db 1381 TTCAGTTCCTTTCTCATCTGGGGCTCTCAAAAATCCAGTTCTGAAATCAACAACGGC 1440
 QY 1596 AGCTGGCACATCAACAGGACTGCCCTAAATTTGCAATGACTCCCTCCAAATGCGTTCTTT 1655
 Db 1441 AGCTGGCACATCAACAGGACTGCCCTAAATTTGCAATGACTCCCTCCAAATGCGTTCTTT 1500
 QY 1656 GCCGGCTTGTACGGACACAAAGTTCAACTGTCGGGTGCGCGGAGCGCATGCGGAGC 1715
 Db 1501 GCCGGCTTGTACGGACACAAAGTTCAACTGTCGGGTGCGCGGAGCGCATGCGGAGC 1560
 QY 1716 TGCGGCCCATTCATCTGTTGCGCCAGGGGTGGGGCCCATCACTATATACTAAGCCTAAC 1775
 Db 1561 TGCGGCCCATTCATCTGTTGCGCCAGGGGTGGGGCCCATCACTATATACTAAGCCTAAC 1620
 QY 1776 AGCTCGGATCAGAGGCTTATTTGCTGGCATTTACGCGCTCGACCGTGTGCTGTCGTC 1835
 Db 1621 AGCTCGGATCAGAGGCTTATTTGCTGGCATTTACGCGCTCGACCGTGTGCTGTCGTC 1680
 QY 1836 CGCTCGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1895
 Db 1681 CGCTCGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 QY 1896 ACCGATCGTTCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1955
 Db 1741 ACCGATCGTTCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 QY 1956 CTCAACAAACGCGTCCGCGACCAAGCAACTGTTTGGCTGTATCATGAGTGAATAGTACT 2015
 Db 1801 CTCAACAAACGCGTCCGCGACCAAGCAACTGTTTGGCTGTATCATGAGTGAATAGTACT 1860
 QY 2016 GGGTTCATTAAGACGCTCGGAGGTCCCGGTGTAACATCGGGGGGTGCGTAAACCGCAC 2075
 Db 1861 GGGTTCATTAAGACGCTCGGAGGTCCCGGTGTAACATCGGGGGGTGCGTAAACCGCAC 1920
 QY 2076 TTGATCTGCCCCACGACTGCTTCCGGAAGCAACCCGAGGCTACTTACACAAAATGTGGC 2135
 Db 1921 TTGATCTGCCCCACGACTGCTTCCGGAAGCAACCCGAGGCTACTTACACAAAATGTGGC 1980
 QY 2136 TCGGGGCCCTGTGTGACCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2195
 Db 1981 TCGGGGCCCTGTGTGACCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
 QY 2196 CCTGTGCTCTCAATTTTCCATCTTAAAGTTAGGATGTAATGTTGGGGGGGTGCGGAGCAC 2255
 Db 2041 CCTGTGCTCTCAATTTTCCATCTTAAAGTTAGGATGTAATGTTGGGGGGGTGCGGAGCAC 2100
 QY 2256 AGGCTCAATCCCGCATGCAATTTGCAATTCGAGGAGAGCGGCTGTAACTTGGAGGACAGGAT 2315
 Db 2101 AGGCTCAATCCCGCATGCAATTTGCAATTCGAGGAGAGCGGCTGTAACTTGGAGGACAGGAT 2160
 QY 2316 AGGTCAAACTAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2375
 Db 2161 AGGTCAAACTAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 QY 2376 TTCACCAACCTAGCCCGCTTATCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2435
 Db 2221 TTCACCAACCTAGCCCGCTTATCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280

2436 GTGCAATACCTGTACGGTGTAGGTCAGCGTTTGTCTCCTTTGCAATCAAAATGGAGTAC 2495
 Db GTGCAATACCTGTACGGTGTAGGTCAGCGTTTGTCTCCTTTGCAATCAAAATGGAGTAC 2340
 2496 ATCTGTTGCTTTTCTCTCTGCGAGACGCGCGGTGTGTCCTGCTGTGTTGGATG 2555
 Db ATCTGTTGCTTTTCTCTCTGCGAGACGCGCGGTGTGTCCTGCTGTGTTGGATG 2400
 2556 CTGCTGATAGCCAGGCTGAGCGCGCTTAGAGAACTTGGTGGTCTCAATGGCGTCC 2615
 Db CTGCTGATAGCCAGGCTGAGCGCGCTTAGAGAACTTGGTGGTCTCAATGGCGTCC 2460
 2616 GTGGCGGAGCGCATGATTTCTCTCTTCTTGTGTCTTCTGCGCGCGCTGGTACAT 2675
 Db GTGGCGGAGCGCATGATTTCTCTCTTCTTGTGTCTTCTGCGCGCGCTGGTACAT 2520
 2676 AAGGCGAGGCTGCTCTGCGCGCGGTATGCTTTTATGGCGTATGGCGCTGCTCCTG 2735
 Db AAGGCGAGGCTGCTCTGCGCGCGGTATGCTTTTATGGCGTATGGCGCTGCTCCTG 2580
 2736 CTCTACTGCGGTTACACACAGGCTTACGCTTGGACCGGAGATGGTGCATCGTGC 2795
 Db CTCTACTGCGGTTACACACAGGCTTACGCTTGGACCGGAGATGGTGCATCGTGC 2640
 2796 GGGGGTCCGCTTCTTGTAGGTCTGGTATCTTGACCTTGTACCACTACAAAGTGT 2855
 Db GGGGGTCCGCTTCTTGTAGGCTTGTATTTCTTGTAGCTTGTACCACTACAAAGTGT 2700
 2856 CTCACTAGGCTCATAGGTGGTGTACAATCTTATACACAGCGCGAGCGCAATGCCA 2915
 Db CTCACTAGGCTCATAGGTGGTGTACAATCTTATACACAGCGCGAGCGCAATGCCA 2760
 2916 GTGTGGTCCCGCTTCAAGTTTGGGAGGCGGATGCTCATCTCTCTCACTGT 2975
 Db GTGTGGTCCCGCTTCAAGTTTGGGAGGCGGATGCTCATCTCTCTCACTGT 2820
 2976 GCGGTTCAACAGAGTTAATTTTGTACATCACCAACTCTGTCTGCGCATCTCGGCCG 3035
 Db GCGGTTCAACAGAGTTAATTTTGTACATCACCAACTCTGTCTGCGCATCTCGGCCG 2880
 3036 CTATGCTGCTCAGGCTGATACAGAGTGGTGTCTGTCGCGCTCAGGCTC 3095
 Db CTATGCTGCTCAGGCTGATACAGAGTGGTGTCTGTCGCGCTCAGGCTC 2940
 3096 ATTGCTGATCAGTGTAGTGGGAAAGTCCGCGGGGTCAATATGTCGCAATGCTTCT 3155
 Db ATTGCTGATCAGTGTAGTGGGAAAGTCCGCGGGGTCAATATGTCGCAATGCTTCT 3000
 3156 ATGAGCTGGCGGCTGACAGGTACGTATGTTTATTAACCATCTTACCCCACTGCGGAC 3215
 Db ATGAGCTGGCGGCTGACAGGTACGTATGTTTATTAACCATCTTACCCCACTGCGGAC 3060
 3216 TGGGCCACCGGCTTACAGAGCTTGGTGGGCTGAGAGCGGCTCTCTCGCC 3275
 Db TGGGCCACCGGCTTACAGAGCTTGGTGGGCTGAGAGCGGCTCTCTCGCC 3120
 3276 ATGAGACCAAGGTATCATCTGCGGAGACACCGCTGCTGTGGGACATCATCTG 3335
 Db ATGAGACCAAGGTATCATCTGCGGAGACACACCGCTGCTGTGGGACATCATCTG 3180
 3336 GGTCTACCGCTCTCGGCGGAGGGGAGAGATATTTTGGGACCGGCTGATGCTC 3395
 Db GGTCTACCGCTCTCGGCGGAGGGGAGAGATATTTTGGGACCGGCTGATGCTC 3240
 3396 GAAGGCAAGGTGCGGCTCTTGGCCCTACGCGCTACTCCCAACAAACCGGGG 3455
 Db GAAGGCAAGGTGCGGCTCTTGGCCCTACGCGCTACTCCCAACAAACCGGGG 3300
 3456 GTACTTGGTTCATCATCTAGCTCAGCGCGGGAACAAGACAGGTCGAAGGGAG 3515
 Db GTACTTGGTTCATCATCTAGCTCAGCGCGGGAACAAGACAGGTCGAAGGGAG 3360

3516 GTTCAAGTGGTTCCTACCGCAACCAATCTTTCTGGGAGCTGCATCAACGCGGTGTC 3575
 Db GTTCAAGTGGTTCCTACCGCAACCAATCTTTCTGGGAGCTGCATCAATGGGCTGTGT 3420
 3576 TGGAGTGTCTACCATGGGCTGGTCCGAAGACCTAGCGGTCGCAAAAGTCCCAATCACC 3635
 Db TGGAGTGTCTACCATGGGCTGGTCCGAAGACCTAGCGGTCGCAAAAGTCCCAATCACC 3480
 3636 CAATATGTACCAATGTAGACTCTGTGGCTGGCAGGCGCCCCCGGGGCGGCG 3695
 Db CAATATGTACCAATGTAGACTCTGTGGCTGGCAGGCGCCCCCGGGGCGGCG 3540
 3696 TCCATGACACCATGACGTGTGGAGCTCGGACCTTACTTGTCAAGAGACATGCTCAT 3755
 Db TCCATGACACCATGACGTGTGGAGCTCGGACCTTACTTGTCAAGAGACATGCTCAT 3600
 3756 GTCATTCGGTGGCGGCGAGCGGACAGCAGGGAAGTCTACTCTCCCGCAGGCCCGTCC 3815
 Db GTCATTCGGTGGCGGCGAGCGGACAGCAGGGAAGTCTACTCTCCCGCAGGCCCGTCC 3660
 3816 TCTTACCTGAAAGGCTCTCTGGGTGGTTCATGCTTTTGGCTTCGGGGAAGTCTGGG 3875
 Db TCTTACCTGAAAGGCTCTCTGGGTGGTTCATGCTTTTGGCTTCGGGGAAGTCTGGG 3720
 3876 GTCTTCCGGCTGCTGTGTCACCGGGGGTGGGAGGCGGTGACCTCATACCCGTT 3935
 Db GTCTTCCGGCTGCTGTGTCACCGGGGGTGGGAGGCGGTGACCTCATACCCGTT 3780
 3936 GAGTCTATGGAACTACCATGGGTCTCGGTCTTCAAGAGCAACTCAACCCCGCGGT 3995
 Db GAGTCTATGGAACTACCATGGGTCTCGGTCTTCAAGAGCAACTCAACCCCGCGGT 3840
 3996 GTACCGGACATTTCCAAAGTGGCAATCTGACAGCTCTTACTGGGAGCGGAGAGCAC 4055
 Db GTACCGGACATTTCCAAAGTGGCAATCTGACAGCTCTTACTGGGAGCGGAGAGCAC 3900
 4056 AAGTCCCGCTGCTGTATGCGCCAAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4115
 Db AAGTCCCGCTGCTGTATGCGCCAAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3960
 4116 GCGCCACCTTGGGCTTGGGCGGTATGTCGCAAGGCAACGCTTACGACCTAACATC 4175
 Db GCGCCACCTTGGGCTTGGGCGGTATGTCGCAAGGCAACGCTTACGACCTAACATC 4020
 3961 GCGCCACCTTGGGCTTGGGCGGTATGTCGCAAGGCAACGCTTACGACCTAACATC 4020
 4176 AGAATGGGGTAAGGACCATTTACCGGCGGTCCATTAAGTACTCCGCTGCTGCAAG 4235
 Db AGAATGGGGTAAGGACCATTTACCGGCGGTCCATTAAGTACTCCGCTGCTGCAAG 4080
 4236 TTCTTGGCGAGCGGTGCTGCTGCGGCGGCTATGACATCATATATGCTGATGAGTGC 4295
 Db TTCTTGGCGAGCGGTGCTGCTGCGGCGGCTATGACATCATATATGCTGATGAGTGC 4140
 4296 CACTCAACTGACTCGACTACCATTTGGGCTCGGCAAGTCTCTGCAAGCGGAGAG 4355
 Db CACTCAACTGACTCGACTACCATTTGGGCTCGGCAAGTCTCTGCAAGCGGAGAG 4200
 4356 GCTGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4415
 Db GCTGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4260
 4416 CACCCCAATATCAGGAATAGGCTGTCCAAATGGAGAGATCCCTTCTATGGCAAA 4475
 Db CACCCCAATATCAGGAATAGGCTGTCCAAATGGAGAGATCCCTTCTATGGCAAA 4320
 4476 GCCATCCCATTCAGGCGCATTAAGGGGGAGGAGCATCTCATTTTCTGCTGCTGCAAG 4535
 Db GCCATCCCATTCAGGCGCATTAAGGGGGAGGAGCATCTCATTTTCTGCTGCTGCAAG 4380
 4536 AATGTCACGAGTCCCGCAAGCTGACAGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTG 4595
 Db AATGTCACGAGTCCCGCAAGCTGACAGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTG 4440
 4596 CCGGGGCTTCTGATGCTCCGTCATACCGCTTATCGGAGAGCTGCTGCTGCTGCTGCTG 4655


```
RESULT 11
AF054258
LOCUS      8780 bp      RNA      linear      VPL 02-MAY-1998
DEFINITION Hepatitis C virus HC-J4, pcv-J4L9 polyprotein gene, partial cds.
ACCESSION AF054258
VERSION    AF054258.1 GI:3098652
KEYWORDS
SOURCE
ORGANISM   Hepatitis C virus
            Hepatitis C virus
            Virus: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE  1 (bases 1 to 8780)
            Yanagi, M., St. Claire, M., Shapiro, M., Emerson, S. U., Purcell, R. H. and
            Bukh, J.
            Transcripts of a chimeric cDNA clone of hepatitis C virus genotype
            1b are infectious in vivo
            Virology 244 (1), 161-172 (1998)
            98240944
            PUMED 9581788
REFERENCE  2 (bases 1 to 8780)
            Yanagi, M. and Bukh, J.
            Direct Submission
            Submitted (18-MAR-1998) Hepatitis Viruses Section, IID, NIAID, NIH,
            ? Center Dr., Building 7, Room 201, Bethesda, MD 20892, USA
            Location/Qualifiers
            1..8780
            /organism="Hepatitis C virus"
            /mol_type="genomic RNA"
            /strain="HC-J4"
            /isolate="pcv-J4L9"
            /db_xref="taxon:11103"
            187..>8780
            /codon_start=1
            /product="polyprotein"
            /protein_id="AAC15731.1"
            /db_xref="GI:3098653"
            /translation="MSTNPKPQKTKRNTNRRPDQVFFGGQIVGGVLLPRRQPLR
            GYRATKASERQPRRRQPIPKARPEGRWAQPGYPWPLYNEGLGWAHLVLSRPG
            SRPQGTDPKRRSRNLGVIDTLCTGCFADLMGYIPLVGAPLGGAAALAGHVRVLED
            GYNAVNLPGGSFSLFLALLSLDIPASAVEVRNVCIGIVHNDNCSNIVYBAAD
            VMHTPGCVPCVREGNSSCWALPTLAARNASVPTTIRHVDLLVGTAAFCSMY
            VGLCSIFLSQLTFFSRHRTADQNCSTPGHVRHAWMDMMNWSPTTALV
            SQLRIPQAVMAGAHVLAGLAYYSVMGNWAKLIVALLFAGVDGHTHTRGV
            GHTSGFTLSFGSASQKQLVNTNGSMHINETALNDSLOTGFFPAALFYAHKFNVS
            GPEPMASCRPIDWFAQGWGPTTYKPNSSDQRPYCHWYAPRPGVVPASVCGPYVC
            FTSPPVVGCTDRSGAPTSWGENETDMLNTRPPQGNWFGCTWNGSTGKTCCGG
            PPCNIGVGNRTLICPTDGERKHEATYTKGSGPWLTERCLVDVYPLRYHPECVNF
            SIFKVMYVGGVHEHRLNACNWRBGRNLEDRESELSPLLSITTEWQILPCAFITL
            PALSTGLTHQHNIQVDVLYGVGSFAFVFAIKWEYILLLFLLADARVCACLWMLL
            IQAENAALENLYLNAASVAGHGLISFLVFFCAAWYIKRGLAPGAAYAFYGVWPLL
            LLLALPPRAYALDREMAASGGAVLVGLVFLTSLPYKVFELRLIIMWQYFTTRAEAH
            MOVVPLNVGRGDALILLTCVHPELIPDITKLALLILGLMVLQAGITRVPYFVR
            AGLIRACMLVAKVAGHYIQWAFMKLGALTGTIVYVNLHLPDRDAHAGRLDLAVAE
            RVVFSAMETKVTWGAADTAACGDIILGLPVSARRKEIILGPAADSLEGCGWLLASIT
            AVSQDRGVLCQITSLTRDKNKGVEGVQVSTATQSLATCINGCWIVIHGASK
            TLGPKRPTITQMTNVLDVGMQAPPGASMTPCSGSDLYLVTRHADVTPVRRG
            DSGSLSPRVSYPILKSGSCGPLLCPSHGVGVFRAAVCTRGVAKAVDIPVSWERT
            MRSPTVDNTPPAVPTQVLAHLPATGSGKSTKVPAAVQAGYKVLVNPVAAVTL
            GFGAYMSKAHGDIPNTRGTITTTGGSTITSYTGKFLADGCGSGGAYDIIICDECHS
            TDSITLIGTGVLDQATAGARLVLAATATPPGSVTVPHNPIEIGLSNNGSIPFYVK
            ALPIEAKGRHLIFCHSKKKDELAALUTGLGNNAVAYIRGLDVSIVPIPGDVVVA
            TDAIMTGFDFDSFSDNCTVQYDFSLDPTFTTETTPQDAYSRRQRRGTGRG
            RSGIYAFVTFGEPSPGMSVLCBQYDAGAWELTPAETSVRIAYLNTPLPQYQ
            DHLFVSESVFTGLTHDAHPLSQTKAGDNFYLVAIQATVCARAQAPPSWDQWKC
            LIRLPTLHGPTLRLVLAGVONEVILTHPIKYIMACSADLEVVTSVWLVGVLA
            ALAAVCLTTGGSVIVGRIILSKPVPVPREVLYQEPDEMERCASQLPFIEQMOLAE
            QFKQALGLQIAKQARAPVVEKSWRAELTFNAKHWNFISGQYLAGLSTLPGN
            PAIASMAFTASITSLTTQHTLLFNILGWAAQLAPPSAASAFYAGIAGAAVGS
            GLKVLVDILAGYAGVAGALVAFKYMGSVEPSTEDLVNLLPAILSPGALVVGVCAA
            ILRRHVGPGEQAWMMNRILAFASRGNHVSPTHYVPESDAAARVITOLLSLTITOLLK
            RLWFQNLNEDCSTPCSGSLRDVMDICTVLTDFKWLQSKLPLRPVPLPSCQSYK
            GYWRGDGIMQVTCPCGAQIAGHVNGSMRLVGRPTCSNTHGTFPFINATYTPGTPSP
            ANYSRALMRVAAEETVEVTRVGDHFYIVTGVITDNDNKCQVFAPEFTEVDGVRUHR
```


QY 3156 ATGAAGCTGGGGCGGCTGACAGGTAAGTACGTTTATACCATCTTACCCACTGGGGAC 3215
 Db 3001 ATGAAGCTGGGGCGGCTGACAGGTAAGTACGTTTATACCATCTTACCCACTGGGGAC 3060
 QY 3216 TGGGGCCACGGGGCGCTACGAGACCTTGGGGTGGGGTAGAGCCCGTCTCTTCTCGGCC 3275
 Db 3061 TGGGGCCACGGGGCGCTACGAGACCTTGGGGTGGGGTAGAGCCCGTCTCTTCTCGGCC 3120
 QY 3276 ATGGAGACCAAGGTATCATCTCTGGGGAGAGACACCGGTGGGGTGGGGACATCATCTTGG 3335
 Db 3121 ATGGAGACCAAGGTATCATCTCTGGGGAGAGACACCGGTGGGGTGGGGACATCATCTTGG 3180
 QY 3336 GGTCTACCGCTCTCGCCGAGAGGAGAGATATTTTGGGACCGGCTCATAGTCTC 3395
 Db 3181 GGTCTACCGCTCTCGCCGAGAGGAGAGATATTTTGGGACCGGCTCATAGTCTC 3240
 QY 3396 GAAAGGCAAGGTTGGCGACTCTTGGGCCCATCACGGCTACTCCCAACAAACGGGGGCG 3455
 Db 3241 GAAAGGCAAGGTTGGCGACTCTTGGGCTCATCACGGCTACTCCCAACAAACGGGGGCG 3300
 QY 3456 GTAATTGGTTGATCATCATCTAGCTTACAGGCGGGAGCAAGACAGGTCAAGGGGAG 3515
 Db 3301 GTAATTGGTTGCTCATCTAGCTTACAGGCGGGAGCAAGACAGGTCAAGGGGAG 3360
 QY 3516 GTTCAAGTGTCTTACCGCAACACATCTTCTGGGACCTGATCAACGGCGGTGTGC 3575
 Db 3361 GTTCAAGTGTCTTACCGCAACACATCTTCTGGGACCTGATCAACGGCGGTGTGC 3420
 QY 3576 TGGACTGTCTACCATGGCGCTGGCTCGAAGACCTAGCCGGTCCAAAGGTCCATCAC 3635
 Db 3421 TGGACTGTCTACCATGGCGCTGGCTCGAAGACCTAGCCGGTCCAAAGGTCCATCAC 3480
 QY 3636 CAAATGTACACCAATGTAGACTGTGACCTGACCTTGGGCGGAGCGGCCCGGGGGCGC 3695
 Db 3481 CAAATGTACACCAATGTAGACTGTGACCTTGGGCGGAGCGGCCCGGGGGCGC 3540
 QY 3696 TCCATGACACCATGACGTGTGCGACCTGCGACTTACTTGTACAGAGATGCTGAT 3755
 Db 3541 TCCATGACACCATGACGTGTGCGACCTTGTGTTGTACAGAGATGCTGAT 3600
 QY 3756 GTCAATTCGGGTGGCGGGAGCGACAGCAGGGGGAAGTCTACTCTCCCGAGGCCGCTC 3815
 Db 3601 GTCAATTCGGGTGGCGGGAGCGACAGCAGGGGGAAGTCTACTCTCCCGAGGCCGCTC 3660
 QY 3816 TCGTACTGAAAGGCTCTCGGGTGTGCTGCTTGGCTTGGGGAGCGCTGCTGGGCG 3875
 Db 3661 TCGTACTGAAAGGCTCTCGGGTGTGCTGCTTGGCTTGGGGAGCGCTGCTGGGCG 3720
 QY 3876 GTCTTCGGGCTGCTGTGTGACCCGGGGGTGCGAAGCGGTGGACTTTCATACCCGTT 3935
 Db 3721 GTCTTCGGGCTGCTGTGTGACCCGGGGGTGCGAAGCGGTGGACTTTCATACCCGTT 3780
 QY 3936 GAGTCTATGGAATACTACCATGGGTCTCCGGTCTTCCAGACAACTCAACCCCGCGGT 3995
 Db 3781 GAGTCTATGGAATACTACCATGGGTCTCCGGTCTTCCAGACAACTCAACCCCGCGGT 3840
 QY 3996 GTACCGGAGACATTCGAAGTGGACATCTGCACTCTCTCTGCGGCGGAGAGAC 4055
 Db 3841 GTACCGGAGACATTCGAAGTGGACATCTGCACTCTCTCTGCGGCGGAGAGAC 3900
 QY 4056 AAAGTCCGGCTGCTGTATGAGCCCAAGGTAAGGTGCTGCTGCTGAACCCGCTCGTT 4115
 Db 3901 AAAGTCCGGCTGCTGTATGAGCCCAAGGTAAGGTGCTGCTGCTGAACCCGCTCGTT 3960
 QY 4116 GCGGCCACTTAGGGTTTGGGGGTATATGTCCAAAGGCAACGGGTATCGACCTTAACATC 4175
 Db 3961 GCGGCCACTTAGGGTTTGGGGGTATATGTCCAAAGGCAACGGGTATCGACCTTAACATC 4020
 QY 4176 AGAACTGGGTAGGACCATTAACCGGGGGGTCCATTAAGTACTCCACTATGGGAAG 4235
 Db 4021 AGAACTGGGTAGGACCATTAACCGGGGGGTCCATTAAGTACTCCACTATGGGAAG 4080
 QY 4236 TTCCTTCCCGAGGTTGCTGTCTGGGGGCGCTTATGACATCAATAATGTGATGATGC 4295

Db 4081 TTCTCTCCGAGCGTGGCTGCTCTGGGGGCGCTTATGACATCAATAATGTGATCAGTGC 4140
 QY 4296 CACTCAACTGATCGACTCAATCTTGGGATCGGACAGTCTTGGACCGAGCGGAGACG 4355
 Db 4141 CATTCAACTGATCGACTCAATCTTGGGATCGGACAGTCTTGGACCGAGCGGAGACG 4200
 QY 4356 GCTGGAGCGGGCTGCTGCTCGCCACGCTACACTCCGGGATCGGTACCGTGCCTA 4415
 Db 4201 GCTGGAGCGGGCTGCTGCTCGCCACGCTACACTCCGGGATCGGTACCGTGCCTA 4260
 QY 4416 CACCCCAATATCGAGGAAATAGGCTGTTCACCAATATGAGAGATCCCTTCTATGGCAAA 4475
 Db 4261 CACCCCAATATCGAGGAAATAGGCTGTTCACCAATATGAGAGATCCCTTCTATGGCAAA 4320
 QY 4476 GCTATCCCATTTGAGGCGCATCAAGGGGGAGGATCTCATTTTCTGCCATTCCTCAAGAG 4535
 Db 4321 GCTATCCCATTTGAGGCGCATCAAGGGGGAGGATCTCATTTTCTGCCATTCCTCAAGAG 4380
 QY 4536 AAATGTGACGAGCTCGCGGCAAAAGCTGACGGCTCGGACTGAACGCTGTAGCATATTAC 4595
 Db 4381 AAATGTGACGAGCTCGCGGCAAAAGCTGACGGCTCGGACTGAACGCTGTAGCATATTAC 4440
 QY 4596 CCGGCGCTTGTATGTCTCGCTCATACCGCTTATCGGAGACGTCGTCTGTCTGGCAACAGAC 4655
 Db 4441 CCGGCGCTTGTATGTCTCGCTCATACCGCTTATCGGAGACGTCGTCTGTCTGGCAACAGAC 4500
 QY 4656 GCTCTATATGACGGGTTTCAACGGGATTTTGAATCAGTGTGATCGACTGAACATACATGTCTC 4715
 Db 4501 GCTCTATATGACGGGTTTCAACGGGATTTTGAATCAGTGTGATCGACTGAACATACATGTCTC 4560
 QY 4716 ACCCAGACAGTCTGACTTTCAGCTTGGATCCCACTTACCAATATGACGACGACCGTGGCC 4775
 Db 4561 ACCCAGACAGTCTGACTTTCAGCTTGGATCCCACTTACCAATATGACGACGACCGTGGCC 4620
 QY 4776 CAAAGCGCGGTGTCTCGCTTCGCAACGGCGAGGTAGAACTGGCAGGGGTAGAGTGGCATC 4835
 Db 4621 CAAAGCGCGGTGTCTCGCTTCGCAACGGCGAGGTAGAACTGGCAGGGGTAGAGTGGCATC 4680
 QY 4836 TACAGGTTTGTACTTCCAGGAGAACCGGCCCTGGGCTGTCGATTCCTGGTCTCGTCTGTG 4895
 Db 4681 TACAGGTTTGTACTTCCAGGAGAACCGGCCCTGGGCTGTCGATTCCTGGTCTCGTCTGTG 4740
 QY 4896 GAGTGTATGACCGGGCTGTCTTGTATGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 4955
 Db 4741 GAGTGTATGACCGGGCTGTCTTGTATGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 4800
 QY 4956 TTGCGGGCTTACCTAAATACACAGGTTGCCGCTTGGCAGGACCACTTGGAGTCTTGG 5015
 Db 4801 TTGCGGGCTTACCTAAATACACAGGTTGCCGCTTGGCAGGACCACTTGGAGTCTTGG 4860
 QY 5016 GAGAGGCTTCCAGGCGCTCACCCACATAGATGCCACTTCTCTGTCACAGACTAAACAG 5075
 Db 4861 GAGAGGCTTCCAGGCGCTCACCCACATAGATGCCACTTCTCTGTCACAGACTAAACAG 4920
 QY 5076 GCAGGAGACAACTTCTTACCTTGTGGCATTCAGCTTACAGTGTGGCCAGGCTCAA 5135
 Db 4921 GCAGGAGACAACTTCTTACCTTGTGGCATTCAGCTTACAGTGTGGCCAGGCTCAA 4980
 QY 5136 GTTCCACTTCCATCTGTTGGGACCAATGTGGAGTGTCTCATTCGGGTGAAACCTTACCTG 5195
 Db 4981 GCTTCCACTTCCATCTGTTGGGACCAATGTGGAGTGTCTCATACGGCTGAAACCTTACCTG 5040
 QY 5196 CACGGCCCAACCCCTGCTGTATAGCTAGGAGCGTCCAAAATGAGGTCTATCTCTCA 5255
 Db 5041 CACGGCCCAACCCCTGCTGTATAGCTAGGAGCGTCCAAAATGAGGTCTATCTCTCA 5100
 QY 5256 CACCCCAATATGAGTGTGATCATGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTG 5315
 Db 5101 CACCCCAATATGAGTGTGATCATGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTG 5160
 QY 5316 ACCTGGGTCTGTAGGCGGAGTCTTTCAGCTTTCGCGCATACTGCTTTCGCGAGCGC 5375

Db 5161 ACCTGGGTGCTGTAGCGGAGTCCCTTGACGCTTGTGGCCGCTACTGCTGACGACAGGC 5220
Qy 5376 AGTGTGCTCATTTGTGGCAGGATCATCTTTCGGGAAGCCAGCTGTCTGTCGACAGG 5435
Db 5221 AGTGTGCTCATTTGTGGCAGGATCATCTTTCGGGAAGCCAGCTGTCTGTCGACAGG 5280
Qy 5436 GAAGTCTCTACAGAGGTTGATGAGATGGAAGAGTGTGCTTCAAACTTCTTACATC 5495
Db 5281 GAAGTCTCTACAGAGGTTGATGAGATGGAAGAGTGTGCTTCAAACTTCTTACATC 5340
Qy 5496 GAGCAGGGAATGAGCTGCGCGAGCAATCAAGCAAAAGGCGCTCGGTTGTGCAAAAG 5555
Db 5341 GAGCAGGGAATGAGCTGCGCGAGCAATCAAGCAAAAGGCGCTCGGTTGTGCAAAAG 5400
Qy 5556 GCCAACGAAGCGGAGGCTGTCTCCGTTGTGAGTCCAAAGTGGCGAGCCCTTGAAG 5615
Db 5401 GCCAACGAAGCGGAGGCTGTCTCCGTTGTGAGTCCAAAGTGGCGAGCCCTTGAAG 5460
Qy 5616 ACCTTCTGGGGAAGCAGATGTGAATTTTATCAGCGGAATACATGACAGGCTTGA 5675
Db 5461 ACCTTCTGGGGAAGCAGATGTGAATTTTATCAGCGGAATACATGACAGGCTTGA 5520
Qy 5676 TCCACTCTGCTGAAACCCCGCATAGCATCATTTGATGCAATTTACAGCTTCTTACAT 5735
Db 5521 TCCACTCTGCTGAAACCCCGCATAGCATCATTTGATGCAATTTACAGCTTCTTACAT 5580
Qy 5736 AGCCGCTCACCCCAACCAACCCCTCTGTTTAAACATTTGGGGGAGTGGCTGCC 5795
Db 5581 AGCCGCTCACCCCAACCAACCCCTCTGTTTAAACATTTGGGGGAGTGGCTGCC 5640
Qy 5796 CAACCTGCTCTCCAGCGCTGGTACAGCTTTCGTGGGCGCGCATCGCGGAGCGGCT 5855
Db 5641 CAACCTGCTCTCCAGCGCTGGTACAGCTTTCGTGGGCGCGCATCGCGGAGCGGCT 5700
Qy 5856 GTTGGCAGCATAGGCTTGGGAAGTCTCTGTGGAATTTTGGGGGAGTGGGAGAGG 5915
Db 5701 GTTGGCAGCATAGGCTTGGGAAGTCTCTGTGGAATTTTGGGGGAGTGGGAGAGG 5760
Qy 5916 GTAGCGCGCCACTGCTGGGCTTTAAGGTATGAGGCGGAGTGGCTTCCACGAGGAC 5975
Db 5761 GTAGCGCGCCACTGCTGGGCTTTAAGGTATGAGGCGGAGTGGCTTCCACGAGGAC 5820
Qy 5976 CTGCTCAACTTACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6035
Db 5821 CTGCTCAACTTACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5880
Qy 6036 GCAGCAATACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6095
Db 5881 GCAGCAATACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5940
Qy 6096 CTGATAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6155
Db 5941 CTGATAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6000
Qy 6156 GAGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6215
Db 6001 GAGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060
Qy 6216 CGGCTTCCACGATGATTAATGAGGACTGCTTACGCGCATGCTTCCGCTGCTGCTGCTGCTGCT 6275
Db 6061 CGGCTTCCACGATGATTAATGAGGACTGCTTACGCGCATGCTTCCGCTGCTGCTGCTGCTGCT 6120
Qy 6276 GATGTTGGGATGGAATGACAGGTTGATGATGATGATGATGATGATGATGATGATGATGAT 6335
Db 6121 GATGTTGGGATGGAATGACAGGTTGATGATGATGATGATGATGATGATGATGATGATGAT 6240
Qy 6336 CTCCTGCGCGGTTACCGGAGTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6395
Db 6181 CTCCTGCGCGGTTACCGGAGTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6240
Qy 6396 TGGCGGGGAGCGCATCATGCAAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6455
Db 6241 TGGCGGGGAGCGCATCATGCAAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6300

Qy 6456 GTCAAAAACGGTTCATAGGATCGTAGGCTTAGAACCTGAGCAACACGTTGGCACGGA 6515
Db 6301 GTCAAAAACGGTTCATAGGATCGTAGGCTTAGAACCTGAGCAACACGTTGGCACGGA 6360
Qy 6516 AGCTTCCCATCAACGATACACACGGAACCTTGCACACCTCCCGCGCCCAACTAT 6575
Db 6361 AGCTTCCCATCAACGATACACACGGAACCTTGCACACCTCCCGCGCCCAACTAT 6420
Qy 6576 TCCAGGCGCTATGGGGGTGCTGCTGAGGAGTACGTTGAGGTTACGCTGTGGGGAT 6635
Db 6421 TCCAGGCGCTATGGGGGTGCTGCTGAGGAGTACGTTGAGGTTACGCTGTGGGGAT 6480
Qy 6636 TTCCACTAGCTACGCGGATGACCACTGACAAACGTAAGTGGCCATGCCAGGTTCCGGCC 6695
Db 6481 TTCCACTAGCTACGCGGATGACCACTGACAAACGTAAGTGGCCATGCCAGGTTCCGGCC 6540
Qy 6696 CCCGAATTTTCCAGGAGTGGATGGAGTGGCTTGCACAGTACGCTCCCGGCTGCAAAA 6755
Db 6541 CCCGAATTTTCCAGGAGTGGATGGAGTGGCTTGCACAGTACGCTCCCGGCTGCAAAA 6600
Qy 6756 CCTTCTTACGAGGAGCGTCAAGTTCAGGTCGGCTCAACCAATACTTGTGCGGTCG 6815
Db 6601 CCTTCTTACGAGGAGCGTCAAGTTCAGGTCGGCTCAACCAATACTTGTGCGGTCG 6660
Qy 6816 CAGCTCCCATGCGAGCCCGAACCGGACGTAACAGTGTCTTACTTCCATGCTCAACCGATCCC 6875
Db 6661 CAGCTCCCATGCGAGCCCGAACCGGACGTAACAGTGTCTTACTTCCATGCTCAACCGATCCC 6720
Qy 6876 TCCCATTTTACAGGAGAGCGGCTAAGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6935
Db 6721 TCCCATTTTACAGGAGAGCGGCTAAGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6780
Qy 6936 GCCAGCTCATAGCTAGCCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6995
Db 6781 GCCAGCTCATAGCTAGCCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6840
Qy 6996 CATGACTCCCGGACGCTGACCTCATCGAGGCAACCTCTTGTGGCGGAGAGATGGGC 7055
Db 6841 CATGACTCCCGGACGCTGACCTCATCGAGGCAACCTCTTGTGGCGGAGAGATGGGC 6900
Qy 7056 GGAACATCATCTCGGCTGAGTTCAGAGATTAAGTGTAAATTTCTGGAATCTTTTCAACCG 7115
Db 6901 GGAACATCATCTCGGCTGAGTTCAGAGATTAAGTGTAAATTTCTGGAATCTTTTCAACCG 6960
Qy 7116 CTTTACGCGAGGCGGATGAGAGGAGATATCCGTCGCGGAGATCTCTGCAAAATCC 7175
Db 6961 CTTTACGCGAGGCGGATGAGAGGAGATATCCGTCGCGGAGATCTCTGCAAAATCC 7020
Qy 7176 AGGAATTTCCCTTACGCGTTCCTCATATGGGCAACCGCCGACTACAATCTTCCACTGCTA 7235
Db 7021 AGGAATTTCCCTTACGCGTTCCTCATATGGGCAACCGCCGACTACAATCTTCCACTGCTA 7080
Qy 7236 GAGTCTGGAAGGACCGGACTACGCTCCTCGGTTGATACAGGATGCCATGCACT 7295
Db 7081 GAGTCTGGAAGGACCGGACTACGCTCCTCGGTTGATACAGGATGCCATGCACT 7140
Qy 7296 ACCAAGGCTTCTTCCCAATACCACTCCACGAGAAAGAGAGCGTTGTCTGACAGAAATCC 7355
Db 7141 ACCAAGGCTTCTTCCCAATACCACTCCACGAGAAAGAGAGCGTTGTCTGACAGAAATCC 7200
Qy 7356 AATGTGTCTTCTGCTTGGCGAGCTGCCACTAAGACCTTTCGCTGCTGCTGCTGCTGCTGCT 7415
Db 7201 AATGTGTCTTCTGCTTGGCGAGCTGCCACTAAGACCTTTCGCTGCTGCTGCTGCTGCTGCT 7260
Qy 7416 GCGTTGTATAGCGGACCGGACCGCTTCTGACCTTGGCTCGAGAGCGTGCACAAA 7475
Db 7261 GCGTTGTATAGCGGACCGGACCGCTTCTGACCTTGGCTCGAGAGCGTGCACAAA 7320
Qy 7476 GGAATCCGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7535
Db 7321 GGAATCCGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7380

```
QY 7536 GATCTCAGCGAGCGGCTCTTGCTTACCGTGAAGTGAAGGCTAGTGAAGTGTGCTGCG 7595
Db 7381 GATCTCAGCGAGCGGCTCTTGCTTACCGTGAAGTGAAGGCTAGTGAAGTGTGCTGCG 7440
QY 7596 TGCTCAATGTCCTATACGTGACAGCGGCGCTGTATCAAGCGCATCGCTCGGAGGAAAGT 7655
Db 7441 TGCTCAATGTCCTATACGTGACAGCGGCGCTGTATCAAGCGCATCGCTCGGAGGAAAGT 7500
QY 7656 AAGCTGCGCCATCAACCGTTCAGCAACTCTTTGCTGCGTCAACCAATGCTTACGCC 7715
Db 7501 AAGCTGCGCCATCAACCGTTCAGCAACTCTTTGCTGCGTCAACCAATGCTTACGCC 7560
QY 7716 ACAACATCCCGACGCAAGCCTCGGACAGAGAGGTCACTTTGACAGATTGCAAGTC 7775
Db 7561 ACAACATCCCGACGCAAGCCTCGGACAGAGAGGTCACTTTGACAGATTGCAAGTC 7620
QY 7776 CTGGATGATCAATTACGGGACGTACTCAAGGAGATGAAGCGAAGGCGGTCCACAGTTAAG 7835
Db 7621 CTGGATGACCAATTACGGGACGTCCTCAAGGAGATGAAGCGAAGGCGGTCCACAGTTAAG 7680
QY 7836 GCTAAGCTTCTATCTATAGAGGCGCTGCAAGCTGAGCGCCGCCACATTCGGCCAAATCC 7895
Db 7681 GCTAAGCTTCTATCTATAGAGGCGCTGCAAGCTGAGCGCCGCCACATTCGGCCAAATCC 7740
QY 7896 AAATTTGGCTATGGGCAAGGAGCTCGGAAACCTATCCAGAGGCGCGTTAAACACATC 7955
Db 7741 AAATTTGGCTATGGGCAAGGAGCTCGGAAACCTATCCAGAGGCGCGTTAAACACATC 7800
QY 7956 CGCTCCGTGAGGAGGACTGCTGAGAGACACTGAACCAACCAATGACACCAACCATCATG 8015
Db 7801 CGCTCCGTGAGGAGGACTGCTGAGAGACACTGAACCAACCAATGACACCAACCATCATG 7860
QY 8016 GCAAAAAGTGAGGTTTCTGCTGCCAACACAGAGAGGAGCGCCGACAGCGCTCGCCTT 8075
Db 7861 GCAAAAAGTGAGGTTTCTGCTGCCAACACAGAGAGGAGCGCCGACAGCGCTCGCCTT 7920
QY 8076 ATCGTATTCGACACCTGGGAGTTCTGTATCGAGAGAGATGGCCCTTTACAGCTGGTC 8135
Db 7921 ATCGTATTCGACACCTGGGAGTTCTGTATCGAGAGAGATGGCCCTTTACAGCTGGTC 7980
QY 8136 TCCACCTTCTCAGGCGGTGATGGCTCTCTATACGATTTCAATCTCCGCCCAACGAG 8195
Db 7981 TCCACCTTCTCAGGCGGTGATGGCTCTCTATACGATTTCAATCTCCGCCCAACGAG 8040
QY 8196 CGGCTGAGTTTCTGGTGAATACCTGGAATCAAGAGAAATCCCTATGGGCTTCTCATAT 8255
Db 8041 CGGCTGAGTTTCTGGTGAATACCTGGAATCAAGAGAAATCCCTATGGGCTTCTCATAT 8100
QY 8256 GACACCGCTGTTTGTACTCAACGGTCACCTGAGAGTGCATTCGTGTGAGGAGTCAATT 8315
Db 8101 GACACCGCTGTTTGTACTCAACGGTCACCTGAGAGTGCATTCGTGTGAGGAGTCAATT 8160
QY 8316 TACCAATGTTGTACTTGGCCCCGAGCCAGACAGCGCTAAGGTGCTCAAGAGCGG 8375
Db 8161 TACCAATGTTGTACTTGGCCCCGAGCCAGACAGCGCTAAGGTGCTCAAGAGCGG 8220
QY 8376 CTTTACATCGGGGTCCCTGACTAATCAAAAGGGCAGAACTCGGCTTATCGCGGTGC 8435
Db 8221 CTTTACATCGGGGTCCCTGACTAATCAAAAGGGCAGAACTCGGCTTATCGCGGTGC 8280
QY 8436 CGGCAAGTGGCGTGTGACGATGCTGGGTAAATACCTTCATCTTACTTTGAAGGCC 8495
Db 8281 CGGCAAGTGGCGTGTGACGATGCTGGGTAAATACCTTCATCTTACTTTGAAGGCC 8340
QY 8496 ACTGCAAGCTGTCAGCTGCAAGCTCCAGGACTGACAGTGTCTGTGAACGGAGAGCAC 8555
Db 8341 ACTGCGGCTGTCAGCTGCAAGCTCCAGGACTGACAGTGTCTGTGAACGGAGAGCAC 8400
QY 8556 CTTGTCTGTTATCTGTGAAGCGCGGAACCCAGAGGATCGCGCGCCCTACAGGCTTC 8615
Db 8401 CTTGTCTGTTATCTGTGAAGCGCGGAACCCAGAGGATCGCGCGCCCTACAGGCTTC 8460
QY 8616 ACGAGGCTATGACTAGGTATTCGCCCCCCCCGGGATCCGCCCAACCAAGTAACGAC 8675
```

```
Db 8461 ACGAGGCTATGACTAGGTATTCGCCCCCCCCGGGATCCGCCCAACCAAGTAACGAC 8520
QY 8676 CTGAGCTGATAACATCATCTTCTCCAAATGTGTGCTAGTCGCGACGATGATCTGGCAAA 8735
Db 8521 CTGAGCTGATAACATCATCTTCTCCAAATGTGTGCTAGTCGCGACGATGATCTGGCAAA 8580
QY 8736 AGGTATACCTACCTCACCCTGACCCACCCACCCCTTTCACGCGGCTGCTGGGAGACA 8795
Db 8581 AGGTATACCTACCTCACCCTGACCCACCCACCCCTTTCACGCGGCTGCTGGGAGACA 8640
QY 8796 GCTAGACACACTCCAACTCTTGCTAGGCAATATCATCATGTATGCGCCACCCCTA 8855
Db 8641 GCTAGACACACTCCAACTCTTGCTAGGCAATATCATCATGTATGCGCCACCCCTA 8700
QY 8856 TGGGCAAGGATGATTCGATGACTCACTTTTCTCCAACTCTTCTAGCTCAAGAGCAACTT 8915
Db 8701 TGGGCAAGGATGATTCGATGACTCACTTTTCTCCAACTCTTCTAGCTCAAGAGCAACTT 8760
QY 8916 GAAAAGCCCTGATTTGCA 8935
Db 8761 GAAAAGCCCTGATTTGCA 8780

RESULT 12
LOCUS AF054251 8779 bp RNA linear VRL 02-MAY-1998
DEFINITION Hepatitis C virus HC-J4, pCV-J4L1, partial sequence.
ACCESSION AF054251
VERSION AF054251.1 GI:3098640
KEYWORDS
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 8779)
AUTHORS Yanagi,M., St Claire,M., Shapiro,N., Emerson,S.U., Purcell,R.H. and
Bukh,J.
TITLE Transcripts of a chimeric cDNA clone of hepatitis C virus genotype
1b are infectious in vivo
JOURNAL Virology 244 (1), 161-172 (1998)
MEDLINE 98240944
PUBMED 9581788
REFERENCE 2 (bases 1 to 8779)
AUTHORS Yanagi,M. and Bukh,J.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1998) Hepatitis Viruses Section, LID, NIAID, NIH,
7 Center Dr., Building 7, Room 201, Bethesda, MD 20892, USA
FEATURES
source
1..8779
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/strain="HC-J4"
/isolate="pCV-J4L1"
/db_xref="taxon:11103"
/notes="defective virus"
misc_feature 187..>8779
/notes="region corresponding to polyprotein gene"
ORIGIN
Query Match 90.5%; Score 8684.8; DB 14; Length 8779;
Best Local Similarity 99.4%; Pred. No 0;
Matches 8727; Conservative 0; Mismatches 52; Indels 1; Gaps 1;
QY 156 CGGTCAGTACCGGAATTGCCAGGACGACCGGTCCTTTCTTGGATCAACCCGCTCAA 215
Db 1 CGGTCAGTACCGGAATTGCCAGGACGACCGGTCCTTTCTTGGATCAACCCGCTCAA 60
QY 216 TGCTCGAATTTGGCGTGGCCCCGCGAGACTGTAGCCGAGTAGTGTGGGTCCGAA 275
Db 61 TGCTCGAATTTGGCGTGGCCCCGCGAGACTGTAGCCGAGTAGTGTGGGTCCGAA 120
QY 276 AGSCCTTGTGGTACTGCCTGATAGGTTGCTTGCAGTGCCTCGGGAGGTCCTCGTAGACG 335
```


Db 121 AGGCTTTGTGTACTGCTGATAGGCTGCTTCGAGTCCCGGGAAGTCTCGTAGACCG 180
QY TGCACCATAGACACGAATCCCTAAACCTCAAGAAACCCAAACCTAACCAACCGCGGC 395
Db 181 TGCACCATAGACACGAATCCCTAAACCTCAAGAAACCCAAACCTAACCAACCGCGGC 240
QY 396 CCACAGAGCGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGAGTTTACCTGTTCGG 455
Db 241 CCACAGAGCGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGAGTTTACCTGTTCGG 300
QY 456 CGCAGGCGCCCGCAGGTTGGGTGTCGCGGACCTAGGAAGGTTCCGAGCGGTCCCAACCT 515
Db 301 CGCAGGCGCCCGCAGGTTGGGTGTCGCGGACCTAGGAAGGTTCCGAGCGGTCCCAACCT 360
QY 516 CGTGAAGGCGCAACCTATCCCAAGGCTCGCCGACCCGAGGCGAGGCTGGGCTCAG 575
Db 361 CGTGAAGGCGCAACCTATCCCAAGGCTCGCCGACCCGAGGCGAGGCTGGGCTCAG 420
QY 576 CCGCGGTACCTTGGGCCCTCTATGGGAATGAGGCTCGGGGTGGGAGGATGGCTCCTG 635
Db 421 CCGCGGTACCTTGGGCCCTCTATGGGAATGAGGCTCGGGGTGGGAGGATGGCTCCTG 480
QY 636 TCACCCCGCGCTCCCGGCTAGTTGGGCGCCGACGAGCCCGCGCTAGGTCGCGTAAC 695
Db 481 TCACCCCGCGCTCCCGGCTAGTTGGGCGCCGACGAGCCCGCGCTAGGTCGCGTAAC 540
QY 696 TTGGGTAAAGTCATCGATACCTTACATGCGGCTTCGCCGATCTCATGGGGTACATTCOG 755
Db 541 TTGGGTAAAGTCATCGATACCTTACATGCGGCTTCGCCGATCTCATGGGGTACATTCOG 600
QY 756 CTGCTCGGCGCCCGCTAGGCGGCGCTGCGAGGCGCTTGGCAACGCTGTCGGGTTCTG 815
Db 601 CTGCTCGGCGCCCGCTAGGCGGCGCTGCGAGGCGCTTGGCGACGCTGTCGGGTTCTG 660
QY 816 GAGCAGCGGCTGAACCTATGCAACAGGAACTTGC CGGTTGCTCTCTCTATCTCTCTC 875
Db 661 GAGCAGCGGCTGAACCTATGCAACAGGAACTTGC CGGTTGCTCTCTCTATCTCTCTC 720
QY 876 TTGGCTGTGCTGCTGCTTTGACCATCCAGCTTCGCGCTTATGAAGTCGCAACGCTGCC 935
Db 721 TTGGCTGTGCTGCTGCTTTGACCATCCAGCTTCGCGCTTATGAAGTCGCAACGCTGCC 780
QY 936 GGGATATACCATCTCAGAACGCTGCTCCAACTCAAGCAATTTGTATAGGAGCGGAC 995
Db 781 GGGATATACCATCTCAGAACGCTGCTCCAACTCAAGCAATTTGTATAGGAGCGGAC 840
QY 996 GTGATCATGCAATCTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
Db 841 GTGATCATGCAATCTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
QY 1056 TGGGTAGCGCTCACTCCGACGCTCGGCGGCGAGAAATGCGAGCTGCCCACTACGACAATA 1115
Db 901 TGGGTAGCGCTCACTCCGACGCTCGGCGGCGAGAAATGCGAGCTGCCCACTACGACAATA 960
QY 1116 CGACGCGCACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1175
Db 961 CGACGCGCACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
QY 1176 GATCTCTGGGATGCTATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
Db 1021 GATCTCTGGGATGCTATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 1236 GACACGTGCGAGACTGCACTGCTCACTATCTCCCGGCGCATGTATAGGTACCGCATG 1295
Db 1081 GCGACAGTGCAGGACTGCACTGCTCACTATCTCCCGGCGCATGTATAGGTACCGCATG 1140
QY 1296 GCTTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1355
Db 1141 GCTTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1356 CGGATCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415

Db 1201 CGGATCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
QY 1416 CTTTGCCTACTATTTCCATGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCT 1475
Db 1261 CTTTGCCTACTATTTCCATGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCT 1320
QY 1476 GCCGGGCTTGACGGGGAGACCCACACGAGGGGAGGCTGCGCGGCCACACCACTCCGGG 1535
Db 1321 GCCGGGCTTGACGGGGAGACCCACACGAGGGGAGGCTGCGCGGCCACACCACTCCGGG 1380
QY 1536 TTTCACTGCTCTTTTCTCATCTGGGGCTCTCAGAAATCCAGCTTCTGAATACCAACGCG 1595
Db 1381 TTTCACTGCTCTTTTCTCATCTGGGGCTCTCAGAAATCCAGCTTCTGAATACCAACGCG 1440
QY 1596 AGCTGCGACATCAACAGGACTGCCCTAAATTCGATGACTCCCTCCAACTGGGTTCTTT 1655
Db 1441 AGCTGCGACATCAACAGGACTGCCCTAAATTCGATGACTCCCTCCAACTGGGTTCTTT 1500
QY 1656 GCCGCGCTGTTTTACGACACACAAAGTTCAACTGCTGCGGGTGC CGGAGCGCATGGCCAGC 1715
Db 1501 GCCGCGCTGTTTTACGACACACAAAGTTCAACTGCTGCGGGTGC CGGAGCGCATGGCCAGC 1560
QY 1716 TGGCGGCCCATTTGACTGCTTCCGCGAGGGGTGGGGCCCCCATCACTATATCAAGCTAATC 1775
Db 1561 TGGCGGCCCATTTGACTGCTTCCGCGAGGGGTGGGGCCCCCATCACTATATCAAGCTAATC 1620
QY 1776 AGCTCGEATCAGAGGCTTTATTTGGGCAATACGCGCTCGACCGCTGCTGCTGCTGCTGCTGCTGCT 1835
Db 1621 AGCTCGEATCAGAGGCTTTATTTGGGCAATACGCGCTCGACCGCTGCTGCTGCTGCTGCTGCTGCT 1680
QY 1836 GGTCTGCGAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1895
Db 1681 GGTCTGCGAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
QY 1896 ACGATCTGTTCCGGTGTCTCTAGCTATAGCTGGGGGAGAAATGAGACACACCTGATGCTC 1955
Db 1741 ACGATCTGTTCCGGTGTCTCTAGCTATAGCTGGGGGAGAAATGAGACACACCTGATGCTC 1800
QY 1956 CTCAACACACGCTCCGCGCAAGCAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2015
Db 1801 CTCAACACACGCTCCGCGCAAGCAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
QY 2016 GGGTTCACTAAGACGCTGCGGAGTCCCGCTGTAACTATCGGGGGGTGCTGCTGCTGCTGCTGCTGCT 2075
Db 1861 GGGTTCACTAAGACGCTGCGGAGGCGCGCGCTGTAACTATCGGGGGGTGCTGCTGCTGCTGCTGCTGCT 1920
QY 2076 TTGATCTGCGCCACGAGCTGCTTCCGAGACACCGCGAGGCTACTTACACAAATGCTGCT 2135
Db 1921 TTGATCTGCGCCACGAGCTGCTTCCGAGACACCGCGAGGCTACTTACACAAATGCTGCT 1980
QY 2136 TCGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2195
Db 1981 TCGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
QY 2196 CCGTGCACCTCTCAATTTTTCATCTTAAAGTTAGGATGATGCTGCGGGGGGTGAGAC 2255
Db 2041 CCGTGCACCTCTCAATTTTTCATCTTAAAGTTAGGATGATGCTGCGGGGGGTGAGAC 2100
QY 2256 AGGCTCAATGCGGATGCAATTTGAGCTCGAGAGCGCTGTAACTTTGGAGGACAGGAT 2315
Db 2101 AGGCTCAATGCGGATGCAATTTGAGCTCGAGAGCGCTGTAACTTTGGAGGACAGGAT 2160
QY 2316 AGGTCAGAACTCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2375
Db 2161 AGGTCAGAACTCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
QY 2376 TTCACCACTTACCGGCTTTTATCCACTGTTGATCCATCTCCATCCATCCATCCATCCATCCATCCAT 2435
Db 2221 TTCACCACTTACCGGCTTTTATCCACTGTTGATCCATCTCCATCCATCCATCCATCCATCCATCCAT 2280
QY 2436 GTGCAATACCTGCTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2495
Db 2281 GTGCAATACCTGCTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340

2496 ATCTGTGCTTTTCCTTCTCTGECAGACGCGCGTGTGTGCTGTGTGATGATG 2555
2341 ATCTGTGCTTTTCCTTCTCTGECAGACGCGCGTGTGTGCTGTGTGATGATG 2400
2556 CTGCTGATAGCCAGGCTGAGCGCTTACAGAACTTGTGTGCTTCAATGCGGCTCC 2615
2401 CTGCTGATAGCCAGGCTGAGCGCTTACAGAACTTGTGTGCTTCAATGCGGCTCC 2460
2616 GTGCGCGAGCGCATGATCTCTCTCTTCTTGTGTCTCTGCGCGCTCGTACAT 2675
2461 GTGCGCGAGCGCATGATCTCTCTCTTCTTGTGTCTCTGCGCGCTCGTACAT 2520
2676 AAGGCGAGGCTGCTCTGCGCGGCGTATGCTTTTATGCGGATGCGGCTGCTCTG 2735
2521 AAGGCGAGGCTGCTCTGCGCGGCGTATGCTTTTATGCGGATGCGGCTGCTCTG 2580
2736 CTCTACTGCGGTACACACAGCTTACGCTTGAACGCGGAGATGCTGATCCTGTC 2795
2581 CTCTACTGCGGTACACACAGCTTACGCTTGAACGCGGAGATGCTGATCCTGTC 2640
2796 GGGGTGCGGTCTCTGTAGCTGTGTATCTTCACTTGTACCATACCTGCTGCTGCT 2855
2641 GGGGTGCGGTCTCTGTAGCTGTGTATCTTCACTTGTACCATACCTGCTGCTGCT 2700
2856 CTCACTAGGCTCATATGCTGTGATCAATACTTTATCAACAGAGCGGACATGCAA 2915
2701 CTCACTAGGCTCATATGCTGTGATCAATACTTTATCAACAGAGCGGACATGCAA 2760
2916 GTGTGGTCCCCCTCAACGTTGCGGAGCGCGGATGCCATCATCTCTCTCACTGT 2975
2761 GTGTGGTCCCCCTCAACGTTGCGGAGCGCGGATGCCATCATCTCTCTCACTGT 2820
2976 GCGGTATCCAGAGTTAATTTTACATACACAACTCTGCTGCCATATCTGCGCGG 3035
2821 GCGGTATCCAGAGTTAATTTTGAATACCACTCTGCTGCCATATCTGCGCGG 2880
3036 CTCATGCTCTCAGGCTGGCATACGAGTGCCTTCTGCGCGCTCAAGGCTC 3095
2881 CTCATGCTCTCAGGCTGGCATACGAGTGCCTTCTGCGCGCTCAAGGCTC 2940
3096 ATCTGTGATGATGTTAGTGGCAAGTTCGCGGAGTGCCTTCTGCGCGCTCAAGGCTC 3155
2941 ATCTGTGATGATGTTAGTGGCAAGTTCGCGGAGTGCCTTCTGCGCGCTCAAGGCTC 3000
3156 ATGAAGTGGGCGCTGACAGGTACGTACGTTTATACCATCTTACCCACTGCGGAC 3215
3001 ATGAAGTGGGCGCTGACAGGTACGTACGTTTATACCATCTTACCCACTGCGGAC 3060
3216 TGGGCGGCGGCGCTACGAGACCTTGGGTGGCGGTAGAGCGCTGCTCTTCTCCGCC 3275
3061 TGGGCGGCGGCGCTACGAGACCTTGGGTGGCGGTAGAGCGCTGCTCTTCTCCGCC 3120
3276 ATGAGACCAAGGTATCATCTTGGGAGGAGACACCGCTGCTGTTGGGACATCATCTG 3335
3121 ATGAGACCAAGGTATCATCTTGGGAGGAGACACCGCTGCTGTTGGGACATCATCTG 3180
3336 GGTCTACCGCTCTCGCGCGGAGGAGAGATATTTTGGGACCGGCTCATAGTCTC 3395
3181 GGTCTACCGCTCTCGCGCGGAGGAGAGATATTTTGGGACCGGCTCATAGTCTC 3240
3396 CAAGGCAAGGCTGCGACTCTTGGCGCATACAGGCTTCTGCGGAGATCATCTG 3455
3241 CAAGGCAAGGCTGCGACTCTTGGCGCATACAGGCTTCTGCGGAGATCATCTG 3300
3456 GATCTTGGTTCATCATCAGCTGCTACAGCGGAGACAGACCGGCTGCTGCTGCTGCT 3515
3301 GATCTTGGTTCATCATCAGCTGCTACAGCGGAGACAGACCGGCTGCTGCTGCTGCT 3360
3516 GTTCAAGTGGTTTCTACCGCAACAATCTTCTGCGGACCTGCTGCTGCTGCTGCTGCT 3575
3361 GTTCAAGTGGTTTCTACCGCAACAATCTTCTGCGGACCTGCTGCTGCTGCTGCTGCT 3420

3576 TGGACTGTCTACCATGGCTGCTCGAAGACCTTAGCGGTTCCAAAGGTCCAAATCACC 3635
3421 TGGACTGTCTACCATGGCTGCTCGAAGACCTTAGCGGTTCCAAAGGTCCAAATCACC 3479
3636 CAAATGATACCAATGTAGACCTTGGCTGCTGCGGAGCGCGCCCGCGGCGCGC 3695
3480 CAAATGATACCAATGTAGACCTTGGCTGCTGCGGAGCGCGCCCGCGGCGCGC 3539
3696 TCCATGACACCAATGTAGCTGCTGCGGAGCTTACTTGTGTACAGAGATGCTGAT 3755
3540 TCCATGACACCAATGTAGCTGCTGCGGAGCTTACTTGTGTACAGAGATGCTGAT 3599
3756 GTCAATTCGCGTGGCGGAGCGGACACAGAGGGAAGTCTACTCTCCCGAGCGCGTC 3815
3600 GTCAATTCGCGTGGCGGAGCGGACACAGAGGGAAGTCTACTCTCCCGAGCGCGTC 3659
3816 TCTACTCTGAAAGGCTCTTGGGTGGTCCATTTGCTTTCGCGGAGCGCTGCGGAG 3875
3660 TCTACTCTGAAAGGCTCTTGGGTGGTCCATTTGCTTTCGCGGAGCGCTGCGGAG 3719
3876 GTCTTCGCGTGTGTGTGCACTCGGCGGTGCGGAGGCGTGGACTTCATACCGGTT 3935
3720 GTCTTCGCGTGTGTGTGCACTCGGCGGTGCGGAGGCGTGGACTTCATACCGGTT 3779
3936 GAGTCTATGAAAGTCTTACCATGCGGCTTCCGCTTTCAGAGCAACTCACTCCCGGCT 3995
3780 GAGTCTATGAAAGTCTTACCATGCGGCTTCCGCTTTCAGAGCAACTCACTCCCGGCT 3839
3996 GTACCGGAGACATTCCTGAGTGGCACTTGCAGCTTCTTGGGAGCGGCGGAGGAGCACC 4055
3840 GTACCGGAGACATTCCTGAGTGGCACTTGCAGCTTCTTGGGAGCGGCGGAGGAGCACC 3899
4056 AAAGTGGCGGCTGTGTGTGCACTGAGGTAAGGTGCTGCTGCTGAAACCGGCT 4115
3900 AAAGTGGCGGCTGTGTGTGCACTGAGGTAAGGTGCTGCTGCTGAAACCGGCT 3955
4116 GCGGCACTTGTAGGCTTGGGCGGTATATGTCGAGGCAAGGTGCTGCTGCTGAAACCGGCT 4175
3960 GCGGCACTTGTAGGCTTGGGCGGTATATGTCGAGGCAAGGTGCTGCTGCTGAAACCGGCT 4019
4176 AGAAGTGGGTAAGGACCATTACCGGCGGCTTCCATGCTGCTTCCAGCTTGGGCAAG 4235
4020 AGAAGTGGGTAAGGACCATTACCGGCGGCTTCCATGCTGCTTCCAGCTTGGGCAAG 4079
4236 TCCCTGCGGCGGCTGTGTGTGCGGCGGCTTATGATGATCATATATGATGATGATG 4295
4080 TCCCTGCGGCGGCTGTGTGTGCGGCGGCTTATGATGATCATATATGATGATGATG 4139
4296 CACTCACTGACTGCTGCTGCTGCGGCGGCTTATGATGATCATATATGATGATGATG 4355
4140 CATTCACTGACTGCTGCTGCTGCGGCGGCTTATGATGATCATATATGATGATGATG 4199
4356 GCTGAGCGGCGCTGCTGCTGCGGCGGCTTATGATGATCATATATGATGATGATG 4415
4200 GTTGGAGCGGCGCTGCTGCTGCGGCGGCTTATGATGATCATATATGATGATGATG 4259
4416 CACCCCAATATCGAGGAATAGGCTGCTCAACATGAGAGATGCTGCTGATGAGGAAA 4475
4260 CACCCCAATATCGAGGAATAGGCTGCTCAACATGAGAGATGCTGCTGATGAGGAAA 4319
4476 GCCATCCCATTTGAGGCGGCTTCAAGGGGCGGAGGCTGCTGCTGCTGCTGCTGCTG 4535
4320 GCCATCCCATTTGAGGCGGCTTCAAGGGGCGGAGGCTGCTGCTGCTGCTGCTGCTG 4379
4536 AATGTGACAGCTGCGGCGGCTTCAAGGGGCGGAGGCTGCTGCTGCTGCTGCTGCTG 4595
4380 AACTGCGAGCTGCGGCGGCTTCAAGGGGCGGAGGCTGCTGCTGCTGCTGCTGCTG 4439
4596 CCGGCGGCTTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4655
4440 CCGGCGGCTTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4499
4656 GCTCTAATGAGGCTTTCACCGGCGGCTTTCAGTGTGCTGCTGCTGCTGCTGCTGCTG 4715

[illegible]

QY 6876 TCCCAATATACAGCAGAGAGCGCTAAGCGTAGCGTGGCTAGAGGCTCTCCGCCCTTTTA 6935
 Db 6720 TCCCAATATACAGCAGAGAGCGCTAAGCGTAGCGTGGCTAGAGGCTCTCCGCCCTTTTA 6779
 QY 6936 GCGAGCTATACAGCAGAGAGCGCTAAGCGTAGCGTGGCTAGAGGCTCTCCGCCCTTTTA 6995
 Db 6780 GCGAGCTATACAGCAGAGAGCGCTAAGCGTAGCGTGGCTAGAGGCTCTCCGCCCTTTTA 6839
 QY 6996 CATGACTCCCGGAGCGCTAGCGTGGCTAGAGGAGCGCTAAGCGTAGCGTGGCTAGAGGAGTGGC 7055
 Db 6840 CATGACTCCCGGAGCGCTAGCGTGGCTAGAGGAGCGCTAAGCGTAGCGTGGCTAGAGGAGTGGT 6899
 QY 7056 GGAACATACATCGCGTGGAGTCAGAGATAGAGTAGTAATCTCGAGCTCTTTTGAACCG 7115
 Db 6900 GGAACATACATCGCGTGGAGTCAGAGATAGAGTAGTAATCTCGAGCTCTTTTGAACCG 6959
 QY 7116 CTTACGCGGAGGAGTGAAGAGGAGATATCGCTGGCGGAGATCGTGGAGAAATCC 7175
 Db 6960 CTTACGCGGAGGAGTGAAGAGGAGATATCGCTGGCGGAGATCGTGGAGAAATCC 7019
 QY 7176 AGGAAGTTCCCTCAGCGTTGCCATATGGGCAAGCGGAGTACCGGATGCCATTTGCCACT 7235
 Db 7020 AGGAAGTTCCCTCAGCGTTGCCATATGGGCAAGCGGAGTACCGGATGCCATTTGCCACT 7079
 QY 7236 GAGTCTCGAAGSACCCGAGTACGCTCCCTCGGTTGTTACCGGATGCCATTTGCCACT 7295
 Db 7080 GAGTCTCGAAGSACCCGAGTACGCTCCCTCGGTTGTTACCGGATGCCATTTGCCACT 7139
 QY 7296 ACCAAGCTCTCTCAATACACCTCCAGGAGAAAGAGAGCGTTGTTGCCAGAAATCC 7355
 Db 7140 ACCAAGCTCTCTCAATACACCTCCAGGAGAAAGAGAGCGTTGTTGCCAGAAATCC 7199
 QY 7356 AATGCTCTCTCGCTTGGCGAGCTCGCATAGAGCTTCGGTACCGGATGCCATTTGCCACT 7415
 Db 7200 AATGCTCTCTCGCTTGGCGAGCTCGCATAGAGCTTCGGTACCGGATGCCATTTGCCACT 7259
 QY 7416 GCGGTTGATAGGCGCAGCGGAGCCGCTTCTGACCTGGCTCCGACGAGCGGTGACAAA 7475
 Db 7260 GCGGTTGATAGGCGCAGCGGAGCCGCTTCTGACCTGGCTCCGACGAGCGGTGACAAA 7319
 QY 7476 GGATCCGAGCTGAGTCTGCTATCTCTCATGCCCCCTTTGAAGGGGAGCGCGGGAGCC 7535
 Db 7320 GGGTCCGAGCTGAGTCTGCTATCTCTCATGCCCCCTTTGAAGGGGAGCGCGGGAGCC 7379
 QY 7536 GATCTCAGCGAGCGGTCTTGGCTACCGTGGAGGAGGCTAGTGGAGTGTCTGCTGC 7595
 Db 7380 GATCTCAGCGAGCGGTCTTGGCTACCGTGGAGGAGGCTAGTGGAGTGTCTGCTGC 7439
 QY 7596 TGCTCAATGCTTATACCTGGAGAGCGCGCTTGATCAGCCATGCGCTCGCGAGGAAAGT 7655
 Db 7440 TGCTCAATGCTTATACCTGGAGAGCGCGCTTGATCAGCCATGCGCTCGCGAGGAAAGT 7499
 QY 7656 AAGCTGCCATCAACCGCTTGAAGCACTCTTGTGCTCATCCAGCAATGGCTGTACGCC 7715
 Db 7500 AAGCTGCCATCAACCGCTTGAAGCACTCTTGTGCTCATCCAGCAATGGCTGTACGCC 7559
 QY 7716 ACAACATCCCGAGCGCAAGCTCCGAGAGAGGAGTCACTTTGACAGATTCAGTGC 7775
 Db 7560 ACAACATCCCGAGCGCAAGCTCCGAGAGAGGAGTCACTTTGACAGATTCAGTGC 7619
 QY 7776 CTGATGATCATTAACCGGAGCTACTCAAGGAGATGAAGGCGAAGGCGTCCACAGTTAAG 7835
 Db 7620 CTGATGATCATTAACCGGAGCTACTCAAGGAGATGAAGGCGAAGGCGTCCACAGTTAAG 7679
 QY 7836 GCTAAGCTCTATATAGAGGAGCGCTGAGCTGAGCGCCCGCAATTCGCGCAATTC 7895
 Db 7680 GCTAAGCTCTATATAGAGGAGCGCTGAGCTGAGCGCCCGCAATTCGCGCAATTC 7739
 QY 7896 AATTTGGCTATGGGCAAGGAGCGTCCGAGACCTATCCAGAGCGCGCTTAAACATC 7955
 Db 7740 AATTTGGCTATGGGCAAGGAGCGTCCGAGACCTATCCAGAGCGCGCTTAAACATC 7799

QY 7956 CGCTCCGTGGGAGGACTTGTGTGAAGAGCACTTGAACCAATTTGACCAACCATCATG 8015
 Db 7800 CGCTCCGTGGGAGGACTTGTGTGAAGAGCACTTGAACCAATTTGACCAACCATCATG 7859
 QY 8016 GCAAAAAGTGAAGTTTCTGGCTCCAAACAGAGAGGAGGCGGCAAGCGCTCGCCTT 8075
 Db 7860 GCAAAAAGTGAAGTTTCTGGCTCCAAACAGAGAGGAGGCGGCAAGCGCTCGCCTT 7919
 QY 8076 ATCGTATTCACAGAGCTGGGAGTTCGTGTATGCGAGAGAGATGGCCCTTTTACAGCTGCTC 8135
 Db 7920 ATCGTATTCACAGAGCTGGGAGTTCGTGTATGCGAGAGAGATGGCCCTTTTACAGCTGCTC 7979
 QY 8136 TCCACCTCTCCCTCAGGCGCTGATGCGCTTCCCTCATACGAGATTTCAATATCCCCCAAGCAG 8195
 Db 7980 TCCACCTCTCTCAGGCGCTGATGCGCTTCCCTCATACGAGATTTCAATATCCCCCAAGCAG 8039
 QY 8196 CGGCTCGAGTTTCTGGTGAATACCTGGAAATCAAAGAAATGCCCTATGGGCTTCTCATAT 8255
 Db 8040 CGGCTCGAGTTTCTGGTGAATACCTGGAAATCAAAGAAATGCCCTATGGGCTTCTCATAT 8099
 QY 8256 GACACCGCTCTTTTGAATCAACGCTCACTGAGAGTGCATTCGTGTGAGGAGTCAATT 8315
 Db 8100 GACACCGCTCTTTTGAATCAACGCTCACTGAGAGTGCATTCGTGTGAGGAGTCAATT 8159
 QY 8316 TACCAATCTTGTGACTTGGCCCGAGGCGCAGAGAGCATAAGGTCTGCTCAGAGAGCGG 8375
 Db 8160 TACCAATCTTGTGACTTGGCCCGAGGCGCAGAGAGCATAAGGTCTGCTCAGAGAGCGG 8219
 QY 8376 CTTTACATCGGCGGTCCCTGACTAACTCAAAAGGCGAGAACTCGGTTATCGCGGTGC 8435
 Db 8220 CTTTACATCGGCGGTCCCTGACTAACTCAAAAGGCGAGAACTCGGTTATCGCGGTGC 8279
 QY 8436 CGCGCAAGTGGCGCTGACGAGCTAGCTCGGTAAATACCTTCACTGTTACTTGAAGGCC 8495
 Db 8280 CGCGCAAGTGGCGCTGACGAGCTAGCTCGGTAAATACCTTCACTGTTACTTGAAGGCC 8339
 QY 8496 ACTGAGCTGTCTGAGCTGCAAAAGTCCAGGACTCCAGGAGTGGCGGCGCTTACGAGCCTC 8555
 Db 8340 ACTGAGCTGTCTGAGCTGCAAAAGTCCAGGACTCCAGGAGTGGCGGCGCTTACGAGCCTC 8399
 QY 8556 CTTTGTGTTATCTGTAAAGCGCGGAAACCCAGAGGAGTGGCGGCGCTTACGAGCCTC 8615
 Db 8400 CTTTGTGTTATCTGTAAAGCGCGGAAACCCAGAGGAGTGGCGGCGCTTACGAGCCTC 8459
 QY 8616 ACGGAGCTGATGACTAGGTATTCGCGCCCGCGGAGTCCGCCCAACAGAGATGAC 8675
 Db 8460 ACGGAGCTGATGACTAGGTATTCGCGCCCGCGGAGTCCGCCCAACAGAGATGAC 8519
 QY 8676 CTGAGCTGATGATCATCATGTTCTCCATGTGTGCTGCGGAGTGGCGGCGCTTACGAGCCTC 8735
 Db 8520 CTGAGCTGATGATCATCATGTTCTCCATGTGTGCTGCGGAGTGGCGGCGCTTACGAGCCTC 8579
 QY 8736 AGGATATCTACCTACCGGTGACCCCAACCCCTTGCAGCGGCTGGTGGAGACA 8795
 Db 8580 AGGATATCTACCTACCGGTGACCCCAACCCCTTGCAGCGGCTGGTGGAGACA 8639
 QY 8796 GCTAGACACACTCCAACTCTGCTAGGAGGAGTGCATGATGCGGCGGCGCTTACGAGCCTC 8855
 Db 8640 GCTAGACACACTCCAACTCTGCTAGGAGGAGTGCATGATGCGGCGGCGCTTACGAGCCTC 8699
 QY 8856 TGGCAAGGATGATCTGAGTCACTTTTCTCCATCTTCTGCTAGCTCAAGAGCACTT 8915
 Db 8700 TGGCAAGGATGATCTGAGTCACTTTTCTCCATCTTCTGCTAGCTCAAGAGCACTT 8759
 QY 8916 GAAAAAGCCCTGGATTGTCA 8935
 Db 8760 GAAAAAGCCCTGGATTGTCA 8779

RESULT 13
 AF054254
 LOCUS
 DEFINITION

Hepatitis C virus HC-J4, pCV-J4L4
 8781 bp
 RNA
 linear
 VRL 02-MAY-1998
 partial cds.

```

ACCESSION AF054254
VERSION AF054254.1 GI:3098645
KEYWORDS Hepatitis C virus
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
REFERENCE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
AUTHORS Hepacivirus.
1 (bases 1 to 8781)
Yanagi,M., St Claire,M., Shapiro,M., Emerson,S.U., Purcell,R.H. and
Bukh,J.
TITLE Transcripts of a chimeric cDNA clone of hepatitis C virus genotype
1b are infectious in vivo
JOURNAL Virology 244 (1), 161-172 (1998)
MEDLINE 98240944
PubMed 9581788
REFERENCE 2 (bases 1 to 8781)
AUTHORS Yanagi,M. and Bukh,J.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1998) Hepatitis Viruses Section, LID, NIAID, NIH,
7 Center Dr., Building 7, Room 201, Bethesda, MD 20892, USA
FEATURES
source
1..8781
Location/Qualifiers
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/strain="HC-J4"
/isolate="pCV-J4L4"
/db_xref="taxon:11103"
188..>8781
/codon_start=1
/product="polyprotein"
/protein_id="AAC15728.1"
/db_xref="GI:3098646"
/translation="MSTNPKPKRTKRNTRRRPQDVKPPGGQIVGGVYLLPRRGPRL
GVRATKTSERSQRRGRQPIPKARRPGRRAWQGPWPPLYGNEGLGWAGWLLSPRG
SRPSWGTDPRRRKSLVDLTGCPADLMGVILVGA PLGGAARALAHGVLEL
GVNATGNLPGCSFISFLALLSLCTIPASAYEVNYSIGIYHVDNCSNIIYBAAD
VIMHTPGCPVPCVRGNSRCWALTPILAAKDAASVPTIIRRHVDLIVTAFCSAM
VGDLCGSIFLVSQITFSPRHETVQDCNCSIYPGHVSCHRMWDMWNMSPTTALVY
SOLLRIPOAVYDVAGHFWGLAGLAYYSVMGNKAVLIVALLFAGVDGETHTTGRVY
GRITTSGLSPSSGASQIGLVNTNGSHWIRNLTALNCNDSLOTGFLAALFYTHFNSP
GCPERMASCRDIPFAOQGWPIITVETPNSSDPQPCWYAPRPGIIPASQVCGPVVC
FTPSPVVGTTDRSGVPTYSWGENETDMLNTRPPQGNWFGCTWNSIGFTKCGG
PPCNIGVGNTEILCPTDCPRKHEATYTKCGSPWLTFRCLVDIPIRLMHPYCTLNF
SIFKRMIVGVGSEHLKNAACNWRGRCNLEDRDSBELSPILLSSTWQILPCATLNF
PALSTGLIHQNIIVDQYLYGVGSFVSFAIKWEYIILLFLLADARVACLWMLL
IAQEALENI.VNLNAASVAGANGILSLPFFCAAWIKGRLPACAAVAFYGVWPLLI
LLALLPRAYALADREMAASCGAVLGLVFLTLSPYKVFUTRLIMLQYFI TRAEAH
MQVWVPLNVRGGDDAIIITLCAVHPELI FDI TKLLALLGLPLMVLQAGITRVPYFVR
AQLIRACMLVRKAGHYVQMAFKLIGALITGVYVNHLLPLRDWARAGURDLAVFE
PVFSAMETKVITWGAUTACGDIILGLPVSARRKKEIFLGPADSLGQGWLLAPIT
AYSQOTRGVLGCVTSITGRDNKNQVEGVQVSTATQSFATCINGVCWTYVHGAGSK
TLAGEKGPITQMYTNVDLIDLVGQAPPGARSMTPCSGSDLVLTWRHADVI PVRRRG
DSRGLSLSPREYSLKSGSGGFLCPSGHVGVFRAAVCTRGVAKVDFI PVESMETT
MRS.PVFTDSSSPPAVPTFOVAHLHAPTGSGSKTKVPAAYAOQYKVLVLPNSVAAL
GFGAMSKAHGIDNIRGVPTITTGSIYSTYKELADGCGSGAVIDIILICEBS
TDSITLIGITVLQDAETAGARLVLAATAPPGSVTPVPHNIEIGLSNNGEIPFYK
AIPLEALKGRRHLLFCHSKKODELAAKTLGLNNAVAYRGLDVSVPITPIDVVVA
RSGTYRFTVPERSPGFMFDSVLCBQDAGCAWHELTPAETSVRILAPNPLPVQC
DHLFVSVFTGLTHIDAFLSOTKQAGNFPYLVAYOATVCAQAQAPPSDQMWQC
LIRLKLPHGPTLLYRLGAVQNEVILTHPIKIMACMSADLVVTSVTLVGVGLA
ALAAVCLITGSVIVUGRIILSGFSAVPDREVLVQEFDEMEBCASQUPYIQGWLAE
QFKRAGLITQATQKAEAPVSEWRALLETFWAKEMNFI SQIYLAGISLTPGN
PALSLMFAFTASITPTQNTLLNLGFWAAQLAPPASAFVAGIADVAGVSI
GLGLVLDI LAGYAGVAGVALVAFKMSGEVPESTDVNLPLAILSPGALVGVVCA
ILRHVGPBGAVQWNNELIAPASGNHVSPTHVVPESDAARVTOILPSTIOLIK
RLHWINEDCSTPCSGWLRVDWDLCTVLDFTKWLQSKLLPLRPLPVPLSCORGTK
GVWGDGIMQNTCPCGAGIAGVNGSMRI VGPRTCSNTHGTFPIINATYTPCTPSP
APNYSRALWRVAABEVTVRWGDFHYVVTGMDTNVRCPCQVPAPBPFTVEVDGVRLR
YAPACRPLLRDEVTFOVGLNQYLGQSLPCEPEPDVYLTSMLTDPHSHTAETAKRL
ARGSPPLSASSASQSLAKATCTTHDSRDPADLLEANLLWQEMGNITRVESN
KWLIDSPFLHAEGDEREISVAAEILRKSRKPSALPI MARPDYNPLLESWKDPDY
VPPVHGCPLPTKAPPIPPRRKRTVVLTESNVSSALAEIATKTFSSGSSAVDSGT

```

```

ATATPDLASDDKGDVSHSMPLEGGPDPLSDGWSVTSBEASBVDVCCMS
YTWIGALITPCAAEESKLPINPLNSILRHNVYATTPRSASIROKKVTPDLQVLD
DHVDVILKEMKASVTKVALLSIEACKLTTPHSASKEFGYGAQDVNLSRAVNH
RSWNEDLLETETPIPTTMAKSEVFCVQFEKGRKPARLIVFPDLGVRVCEKALDV
VSTLPQAVVSSYGFQSPQRVEFLVNTWKSKPMGFSDYTRCPDSTVTSDDIRV
EESYQCCDLAPARQAIRLSLIRLYIGGLTNSKGNCGYRRCRASGLVTTSCGNFL
TCYKATACAAKALQDCTMLVNGDDILVVCESAGTQEDAAALRAFEAMTRYSAFPG
DPQPEYDLELITSCSNSVSAHDASGRVYVLTRODPTPLARAAWETARHTTINSWL
GNLIMYAPTWASMLTWFFSILLAQEQLEKALDC"
variation
/notes="difference between this isolate and pCV-J4L6S"
/replace=""
Query Match 90.2%; Score 8658.6; DB 14; Length 8781;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 871; Conservative 0; Mismatches 69; Indels 1; Gaps 1;
Qy 156 CCGGTGAGTACACCGGAATTGCCAGGACGACCGGGTCTCTTTCTTGGATC-AAACCGGTCA 214
Db 1 CCGGTGAGTACACCGGAATTGCCAGGACGACCGGGTCTCTTTCTTGGATCAAAACCGGTCA 60
Qy 215 ATGCTCGAGATTGGCGTGCCTCCCGCGAGACTGTAGCCGAGTAGTGTGGTCCGGA 274
Db 61 ATGCTCGAGATTGGCGTGCCTCCCGCGAGACTGTAGCCGAGTAGTGTGGTCCGGA 120
Qy 275 AAGCCCTTGTGGTACTGCTGATAGGTGCTTGCAGTGCCTCCCGGAGGTCTCTGTAGACC 334
Db 121 AAGCCCTTGTGGTACTGCTGATAGGTGCTTGCAGTGCCTCCCGGAGGTCTCTGTAGACC 180
Qy 335 GTGCACCATGAGCACGAATCCTAAACCTCAAGAAAACCAAAAGTAAACCAACCCCG 394
Db 181 GTGCACCATGAGCACGAATCCTAAACCTCAAGAAAACCAAAAGTAAACCAACCCCG 240
Qy 395 CCCACAGACGTCAAGTTCCTCCGGCGGTGTGTAGATCGTTCGTGGAGTTCACCTGTGCC 454
Db 241 CCCACAGACGTCAAGTTCCTCCGGCGGTGTGTAGATCGTTCGTGGAGTTCACCTGTGCC 300
Qy 455 GCGCAGGGGCCCCAGGTGTGGTGTGCGCGGACTAGGAAGGCTTCCGAGCGGTGCGCAACC 514
Db 301 GCGCAGGGGCCCCAGGTGTGGTGTGCGCGGACTAGGAAGGCTTCCGAGCGGTGCGCAACC 360
Qy 515 TCGTGAAGGCGCAACACTATCCAAAGGTTCGCGGACCGAGGCGAGGCGCTGGGCTCA 574
Db 361 TCGTGAAGGCGCAACACTATCCAAAGGTTCGCGGACCGAGGCGAGGCGCTGGGCTCA 420
Qy 575 GCGCGGTACCTTGGCCCTCTATGCAATAGGGCTTGGGCTGGGCGAGATGGCTCCT 634
Db 421 GCGCGGTACCTTGGCCCTCTATGCAATAGGGCTTGGGCTGGGCGAGATGGCTCCT 480
Qy 635 GTCAACCCCGGGCTCCCGGCTAGTTGGGCGCCACAGGACCCCGGCTAGGTGCGGTAA 694
Db 481 GTCAACCCCGGGCTCCCGGCTAGTTGGGCGCCACAGGACCCCGGCTAGGTGCGGTAA 540
Qy 695 CTTGGTAAAGTCAATGATACCTTATCATCGGCTTCGCGATCTCATGGGTACATTC 754
Db 541 CTTGGTAAAGTCAATGATACCTTATCATCGGCTTCGCGATCTCATGGGTACATTC 600
Qy 755 GCTCGTCGGGCCCCCTAGGGGCGCTGCGAGGCTTGGCACGCGTCTCCGGGTCT 814
Db 601 GCTCGTCGGGCCCCCTAGGGGCGCTGCGAGGCTTGGCACGCGTCTCCGGGTCT 660
Qy 815 GGAGACGGGCTGAACTATCAACAGGAACTTGGCCGGTGTGCTCTCTCTATCTTCCT 874
Db 661 GGAGACGGGCTGAACTATCAACAGGAACTTGGCCGGTGTGCTCTCTCTATCTTCCT 720
Qy 875 CTTGGCTCTGCTGTCTGTGTGACATCCAGTTCGGTTCGATTAAGTGCAGCGATGTC 934
Db 721 CTTGGCTCTGCTGTCTGTGTGACATCCAGTTCGGTTCGATTAAGTGCAGCGATGTC 780
Qy 935 CGGATATACATGTACAGAACACTGCTCCAACTCAGCAATGTGTATCAGCAGCGGA 994
Db 781 CGGATATACATGTACAGAACACTGCTCCAACTCAGCAATGTGTATCAGCAGCGGA 840

```

QY 995 CTTGATCATGATCACTACCTCCGGGTGCGTCCCTGTGTTCAGGAGGTAAACAGCTCCCGTTG 1054
DB 841 CTTGATCATGATCACTACCTCCGGGTGCGTCCCTGTGTTCAGGAGGTAAACAGCTCCCGTTG 900
QY 1055 CTTGGTAGCGGTCACTCCCAACGCTCCGCGCCAGGAATGCCAGCGTCCCACTACGACAAAT 1114
DB 901 CTTGGTAGCGGTCACTCCCAACGCTCCGCGCCAGGAATGCCAGCGTCCCACTACGACAAAT 960
QY 1115 ACAGCCACAGCTGACCTGCTCGTTGGGACGGCTGCTTCTGCTCCGCTATGATGAGTGG 1174
DB 961 ACAGCCACAGCTGACCTGCTCGTTGGGACGGCTGCTTCTGCTCCGCTATGATGAGTGG 1020
QY 1175 GGAATCTCGCGGANTATTTTCTCGTCTCCAGCTGTTACCTTCTGCTCCGCTATGATGAGTGG 1234
DB 1021 GGAATCTCTGCGGANTATTTTCTCGTCTCCAGCTGTTACCTTCTGCTCCGCTATGATGAGTGG 1080
QY 1235 TGAGACAGTGCAGGACTGCAACTGCTCAATCTATCTCCGCGCCATGATCAGGTCAACGAT 1294
DB 1081 TGAGACAGTGCAGGACTGCAACTGCTCAATCTATCTCCGCGCCATGATCAGGTCAACGAT 1140
QY 1295 GGCCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1354
DB 1141 GGCCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1355 CCGGATCCCAACAGCTGCTGCTGACATGCTGGGCGGCGCCACATGGGGAATGCTGGCGGG 1414
DB 1201 CCGGATCCCAACAGCTGCTGCTGACATGCTGGGCGGCGCCACATGGGGAATGCTGGCGGG 1260
QY 1415 CTTTGCTACTATTTCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1474
DB 1261 CTTTGCTACTATTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 1475 TGCCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1534
DB 1321 TGCCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
QY 1535 GTTCACTGCTCTTTCTCATCTGGGCGGCTCAGAAATCCAGCTTGTGAATACCAACGG 1594
DB 1381 GTTCACTGCTCTTTCTCATCTGGGCGGCTCAGAAATCCAGCTTGTGAATACCAACGG 1440
QY 1595 CAGCTGGGACATCAACAGGACTGCTTCAATGATGATGATGATGATGATGATGATGATGATGAT 1654
DB 1441 CAGCTGGGACATCAACAGGACTGCTTCAATGATGATGATGATGATGATGATGATGATGATGAT 1500
QY 1655 TGCCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1714
DB 1501 TGCCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
QY 1715 CTTCCGCGCCCATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1774
DB 1561 CTTCCGCGCCCATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
QY 1775 CAGCTGGGATCAGAGGCTTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1834
DB 1621 CAGCTGGGATCAGAGGCTTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
QY 1835 CCGCTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1894
DB 1681 CCGCTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
QY 1895 CACCGATCGTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1954
DB 1741 CACCGATCGTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
QY 1955 CTTCAACACAGGCTCCGCCACAGGCAATGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2014
DB 1801 CTTCAACACAGGCTCCGCCACAGGCAATGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
QY 2015 TGGGTTCACTAAGACGTCGAGGTCCCGCTGTAATACATCGGGGGGTGCTGTAACCGGAC 2074
DB 1861 TGGGTTCACTAAGACGTCGAGGTCCCGCTGTAATACATCGGGGGGTGCTGTAACCGGAC 1920

QY 2075 CTTGATCTGCCCCACGACTGCTTCCGAGACACCCCGAGGCTACTTACACAAAATGCG 2134
DB 1921 CTTGATCTGCCCCACGACTGCTTCCGAGACACCCCGAGGCTACTTACACAAAATGCG 1980
QY 2135 CTTGGGGCCCTGGTTGACACCTAGGTGCTTAGTAGCTACCCATACAGGCTTTGGCACTA 2194
DB 1981 CTTGGGGCCCTGGTTGACACCTAGGTGCTTAGTAGCTACCCATACAGGCTTTGGCACTA 2040
QY 2195 CCCCTGACATCTCAATTTTCCATCTTTAAAGGTTAGGATGATGTTGGGGGCGGTGGAGCA 2254
DB 2041 CCCCTGACACCTCAATTTTCCATCTTTAAAGGTTAGGATGATGTTGGGGGCGGTGGAGCA 2100
QY 2255 CAGGCTCAATGCGCGCATGCAATGGACTCGAGAGAGCGCTGTAACTTGGAGGACAGGGA 2314
DB 2101 CAGGCTCAATGCGCGCATGCAATGGACTCGAGAGAGCGCTGTAACTTGGAGGACAGGGA 2160
QY 2315 TAGGTCAGAACTCAGCCCGCTGCTGCTCAACAGAGTGGCAGATACGCTCCCTGTGC 2374
DB 2161 TAGGTCAGAACTCAGCCCGCTGCTGCTCAACAGAGTGGCAGATACGCTCCCTGTGC 2220
QY 2375 TTTTCAACACCTTACCGGCTTTATCCACTGGTTTGAATCCATCTCCATCAGAAATCGTGA 2434
DB 2221 CTTTCAACACCTTACCGGCTTTATCCACTGGTTTGAATCCATCTCCATCAGAAATCGTGA 2280
QY 2435 CTTTCAACACCTTACCGGCTTTATCCACTGGTTTGAATCCATCTCCATCAGAAATCGTGA 2494
DB 2281 CTTTCAACACCTTACCGGCTTTATCCACTGGTTTGAATCCATCTCCATCAGAAATCGTGA 2340
QY 2495 CATCTGTTGCTTTTCTTCTCTGTCAGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2554
DB 2341 CATCTGTTGCTTTTCTTCTCTGTCAGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
QY 2555 GCTGCTGATAGCCAGGCTGAGCGCGCTTACAGAACTTGGTGGTCTCAATCGCGCTC 2614
DB 2401 GCTGCTGATAGCCAGGCTGAGCGCGCTTACAGAACTTGGTGGTCTCAATCGCGCTC 2460
QY 2615 CTTGCGCGGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2674
DB 2461 CTTGCGCGGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
QY 2675 TAAAGGCGAGGCTGCTGCTGCGGCGGTATGCTTTTATGCGCTATGCGCTGCTGCTGCTGCT 2734
DB 2521 TAAAGGCGAGGCTGCTGCTGCGGCGGTATGCTTTTATGCGCTATGCGCTGCTGCTGCTGCT 2580
QY 2735 GCTCCTACTGCGGTTTACACACAGAGCTTACGCTTGAACCGGAGATGGCTGCTGCTGCTGCT 2794
DB 2581 GCTCCTACTGCGGTTTACACACAGAGCTTACGCTTGAACCGGAGATGGCTGCTGCTGCTGCTGCT 2640
QY 2795 CCGGGGTCGGTTCCTTGTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2854
DB 2641 CCGGGGTCGGTTCCTTGTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
QY 2855 TCTCACTAGGCTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2914
DB 2701 TCTCACTAGGCTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760
QY 2915 AGTGTGGTCCCGCTTACAGCTTGGGAGGCGCGGATGCCATCATCTCTCTCAAGTGT 2974
DB 2761 AGTGTGGTCCCGCTTACAGCTTGGGAGGCGCGGATGCCATCATCTCTCTCAAGTGT 2820
QY 2975 TGGGTTTCACTCAGAGTTAAATTTTGAATCACCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3034
DB 2821 TGGGTTTCACTCAGAGTTAAATTTTGAATCACCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880
QY 3035 GCTCATGCTCTCCAGGCTGGCATAACAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3094
DB 2881 GCTCATGCTCTCCAGGCTGGCATAACAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940
QY 3095 CATTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3154
DB 2941 CATTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000
QY 3155 CATGAAGCTGGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3214

Db 3001 CATGAAGCTGGCGCGCTGACAGGTAACGTACGTTTATTAACCATCTTACCCCACTCGCGGA 3060
QY 3215 CTGGCCCAACCGCGGCTTACAGACCTTTCGCGTGGCGGTAGAGCCGTCGTCTTCTCCGC 3274
Db 3061 CTGGCCCAACCGCGGCTTACAGACCTTTCGCGTGGCGGTAGAGCCGTCGTCTTCTCCGC 3120
QY 3275 CATGGAGACCAAGGTATCATCTGGGGAGAGACACCGCTCGGTGGGGACATCATCTT 3334
Db 3121 CATGGAGACCAAGGTATCATCTGGGGAGAGACACCGCTCGGTGGGGACATCATCTT 3180
QY 3335 GGGTCTACCCGCTCTCCGCCCGGAAGGGAGAGATATTTTGGGACCGGCTGATGCTCT 3394
Db 3181 GGGTCTACCCGCTCTCCGCCCGGAAGGGAGAGATATTTTGGGACCGGCTGATGCTCT 3240
QY 3395 CGAAGGGCAAGGGTGGGACTCCITGGGCCCATCACCGGCCCTACCCCAACAAACCGCGGG 3454
Db 3241 CGAAGGGCAAGGGTGGGACTCCITGGGCCCATCACCGGCCCTACCCCAACAAACCGCGGG 3300
QY 3455 CGTACTTGGTTGCATCATCATAGCTTCAAGGCCGGGACAAAGACAGGTCGAAGGGGA 3514
Db 3301 CGTACTTGGTTGCCTATCATAGCTTCAAGGCCGGGACAAAGACAGGTCGAAGGGGA 3360
QY 3515 GGTTCAGAGTGGTTTCTACCGCAACACATCTTCTCGCGACCTGATCAACGGCGTGTG 3574
Db 3361 GGTTCAGAGTGGTTTCTACCGCAACACATCTTCTCGCGACCTGATCAACGGCGTGTG 3420
QY 3575 CTGGACTGTCTACATGGCGCTGCTCGAAGACCTAGCGGTCCTCAAAAGGTCCAATCAC 3634
Db 3421 CTGGACTGTCTACATGGCGCTGCTCAAAAGACCTAGCGGTCCTCAAAAGGTCCAATCAC 3480
QY 3635 CCAATGTACACCAATGTAGACCTGAGCTCGTGGCTGACAGCGCCCGCGGGGGCG 3694
Db 3481 CCAATGTACACCAATGTAGACCTGAGCTCGTGGCTGACAGCGCCCGCGGGGGCG 3540
QY 3695 CTCATGACACCATGAGCTGTGCGAGCTCGGACTTCTTACTTGTACAGAGACATGCTGA 3754
Db 3541 CTCATGACACCATGAGCTGTGCGAGCTCGGACTTCTTACTTGTACAGAGACATGCTGA 3600
QY 3755 TGTCAATTCGGTGGCGGAGCGGACAGCAGGGGAGTCTACTCTCCGCCAGGCGCCGT 3814
Db 3601 TGTCAATTCGGTGGCGGAGCGGACAGCAGGGGAGTCTACTCTCCGCCAGGCGCCGT 3660
QY 3815 CTCCTACTGAAAGGCTCCCTCGGTGCTCCATGCTTGTGCTTCCGGGACAGCTCGTGG 3874
Db 3661 CTCCTACTGAAAGGCTCCCTCGGTGCTTGTGCTTGTGCTTCCGGGACAGCTCGTGG 3720
QY 3875 CGTCTTCGGGCTGTGTGTGACCCCGGGGGTTCGGAAGCGGTGGAATTCATACCCGT 3934
Db 3721 CGTCTTCGGGCTGTGTGTGACCCCGGGGGTTCGGAAGCGGTGGAATTCATACCCGT 3780
QY 3935 TGAAGTCTATGGAACCTACATGGGTCTCGGTCTTCAGAGCACTCAACCCCGCGG 3994
Db 3781 TGAAGTCTATGGAACCTACATGGGTCTCGGTCTTCAGAGCACTCAACCCCGCGG 3840
QY 3995 TGTACCGGACATTCGAAGTGGCAATCTGACGCTCTTACTGSCAGGCGGAAGAGCAC 4054
Db 3841 TGTACCGGACATTCGAAGTGGCAATCTGACGCTCTTACTGSCAGGCGGAAGAGCAC 3900
QY 4055 CAAGTCCCGGCTGTGTATGACGCCCAAGGTACAGGTGCTGCTGCTGAAACCCGTCCGT 4114
Db 3901 CAAGTCCCGGCTGTGTATGACGCCCAAGGTACAGGTGCTGCTGCTGAAACCCGTCCGT 3960
QY 4115 TCCCGCACCTTAGGTTTGGGGCGTATATGTCGAAGGCACAGGTATCGACCCCTAAT 4174
Db 3961 TCCCGCACCTTAGGTTTGGGGCGTATATGTCGAAGGCACAGGTATCGACCCCTAAT 4020
QY 4175 CAGAACTGGGGTAAGGACCAATTAACAGCGCGGCTCCATTAGCTACTCCACCTATGGCAA 4234
Db 4021 CAGAACTGGGGTAAGGACCAATTAACAGCGCGGCTCCATTAGCTACTCCACCTATGGCAA 4080
QY 4235 GTTCTCTCCGACGCTGTGTCTTGGGGGGCGCTATGACATCAATATGTGATGATG 4294

Db 4081 GTTCTCTCCGACGCGTGGCTCTCCGGGGCGCTATGACATCATATATGTGATGATG 4140
QY 4295 CCATCTCAATGACTCGACTACCATCTTTGGGCATCGGCACAGTCTCTGGACCAAGCGGAGC 4354
Db 4141 CCATCTCAATGACTCGACTACCATCTTTGGGCATCGGCACAGTCTCTGGACCAAGCGGAGC 4200
QY 4355 GGTCTGAGCGCGGCTCGTCTGTCTGCGCACCGCTACACCTCCGGGATCGGTTACCGTGCC 4414
Db 4201 GGTCTGAGCGCGGCTCGTCTGTCTGCGCACCGCTACACCTCCGGGATCGGTTACCGTGCC 4260
QY 4415 ACACCCCAATATCGAGAAATAGCCCTGTCCAAACAATGGAGAGATCCCTTCTATGCCAA 4474
Db 4261 ACACCCCAATATCGAGAAATAGCCCTGTCCAAACAATGGAGAGATCCCTTCTATGCCAA 4320
QY 4475 AGCATCTCCCATTTGAGGCCATCAAGGGGGGAGGCAATCTCATTTTCTGTGCATTTCCAGAA 4534
Db 4321 AGCATCTCCCATTTGAGGCCATCAAGGGGGGAGGCAATCTCATTTTCTGTGCATTTCCAGAA 4380
QY 4535 GAAATGTGACGAGCTCGCCGCAAAAGTGCACAGCCCTCGGACTGAAAGCTGTAGCATATTA 4594
Db 4381 GAAATGTGACGAGCTCGCCGCAAAAGTGCACAGCCCTCGGACTGAAAGCTGTAGCATATTA 4440
QY 4595 CCGGGCCTTGATGTGTCCGTTCATACCGCTATCGGAGACGTCGTTGTGTCGTCGCAACAGA 4654
Db 4441 CCGGGCCTTGATGTGTCCGTTCATACCGCTATCGGAGACGTCGTTGTGTCGTCGCAACAGA 4500
QY 4655 CGCTCTAATGACGGGTTTCAACGGCGCATTTTGAATCCCACTTCAACATTGAGACCAACCGTGCC 4714
Db 4501 CGCTCTAATGACGGGTTTCAACGGCGCATTTTGAATCCCACTTCAACATTGAGACCAACCGTGCC 4560
QY 4715 CACCCAGACGTCGACTTCAGCTTGGATCCCACTTCAACATTGAGACCAACCGTGCC 4774
Db 4561 CACTCAGACAGTCGACTTCAGCTTGGATCCCACTTCAACATTGAGACCAACCGTGCC 4620
QY 4775 CCAAGACGGGTTGTCGCGTCGCAACGGCGAGTGAACCTGCGAGGTAGGAGTGGCAT 4834
Db 4621 CCAAGACGGGTTGTCGCGTCGCAACGGCGAGTGAACCTGCGAGGTAGGAGTGGCAT 4680
QY 4835 CTACAGTTTGTGACTCCAGGAGACGGCCCTCGGCACTGTTGATTTCTTCGGTCTGTG 4894
Db 4681 CTACAGTTTGTGACTCCAGGAGACGGCCCTCGGCACTGTTGATTTCTTCGGTCTGTG 4740
QY 4895 TGAGTCTATGACGGGGCTGTGCTTGTGATGAGTCAACCGCGCTGAGACCTCCGTTAG 4954
Db 4741 TGAGTCTATGACGGGGCTGTGCTTGTGATGAGTCAACCGCGCTGAGACCTCCGTTAG 4800
QY 4955 GTTGGGGTTACCTAAATACACAGAGGTTGCCCTCTGCGAGGACCAATCTGGAATTCG 5014
Db 4801 GTTGGGGTTACCTAAATACACAGGATTTGCCCTCTGCGAGGACCAATCTGGAATTCG 4860
QY 5015 GGAGAGCGTCTTTCAGAGGCTTCACCCACATAGATGCCACTTCTGTGCCAGACTAAACA 5074
Db 4861 GGAGAGCGTCTTTCAGAGGCTTCACCCACATAGATGCCACTTCTGTGCCAGACTAAACA 4920
QY 5075 GGCAGGACAGAACTTCTTACCTGGTGGCATATCAAGCTACAGTGTGCGCCAGGCTCA 5134
Db 4921 GGCAGGACAGAACTTCTTACCTGGTGGCATATCAAGCTACAGTGTGCGCCAGGCTCA 4980
QY 5135 AGCTCCACTCCATGCTGGGACCAAAATGTGAAAGTGTCTATAGGCTGAAACCTACTACT 5194
Db 4981 AGCTCCACTCCATGCTGGGACCAAAATGTGAAAGTGTCTATAGGCTGAAACCTACTACT 5040
QY 5195 GCACGGGCCAACCCCTGTGTATAGGCTAGGAGCGCTCCAAATAGGCTCACTCCCTAC 5254
Db 5041 GCACGGGCCAACCCCTGTGTATAGGCTAGGAGCGCTCCAAATAGGCTCACTCCCTAC 5100
QY 5255 ACACCCCAATTAATCATCATGCGCATGCTGCGCTGACCTGAGGTGCTCACTAG 5314
Db 5101 ACACCCCAATTAATCATCATGCGCATGCTGCGCTGACCTGAGGTGCTCACTAG 5160
QY 5315 CACTGTGGTGTGTGAGCGGAGTCTTGTGAGCTTGTGCGCATACTGCGCTGAGCAGG 5374
Db 5161 CACTGTGGTGTGTGAGCGGAGTCTTGTGAGCTTGTGCGCATACTGCGCTGAGCAGG 5220

QY	5375	CAGTGTGGTCATTGTGGCAGGATCATCTGTGCCGGGAAGCAGCTGTCTGTTCCGGACAG	53434
DB	5221	CAGTGTGTGTCATTGTGGCAGGATCATCTGTGCCGGGAAGCAGCTGTCTGTTCCGGACAG	52800
QY	5435	GGAAGTCTCTTACCGAGGAGTTCGATCAGATGAAAGAGTGTGCCTCAACAATTCCTTACAT	5494
DB	5281	GGAAGTCTCTTACCGAGGAGTTCGATCAGATGAAAGAGTGTGCCTCAACAATTCCTTACAT	5340
QY	5495	CGACGAGGAATGCAGCTTCGCGCAGCAATTCAGCAAAAGCGCTCGGTTGTGGAAAC	5554
DB	5341	CGACGAGGAATGCAGCTTCGCGCAGCAATTCAGCAAAAGCGCTCGGTTGTGGAAAC	5400
QY	5555	GGCCACCAGCAAGCGGAGGCTGTGCTCCCGTGTGGAGTCCAAAGTGGCGAGCCCTTGA	5614
DB	5401	GGCCACCAGCAAGCGGAGGCTGTGCTCCCGTGTGGAGTCCAAAGTGGCGAGCCCTTGA	5460
QY	5615	GACCTTCTGGGGGAAGCACTGTGGAAATTCATCAGCGGAATACAGTACCTAGCAGCCTT	5674
DB	5461	GACCTTCTGGGGGAAGCACTGTGGAAATTCATCAGCGGAATACAGTACCTAGCAGCCTT	5520
QY	5675	ATCCACTCTGCTGGAAACCCCGGATAGCATCATTCATGGCGGAATACAGTACCTATCAC	5734
DB	5521	ATCCACTCTGCTGGAAACCCCGGATAGCATCATTCATGGCGGAATACAGTACCTATCAC	5580
QY	5735	TAGCCGCTCACCACCCAAAACACCCCTCTGTTTAAATCTTTGGGGGATGGGTGGCTGC	5794
DB	5681	TAGCCGCTCACCACCCAAAACACCCCTCTGTTTAAATCTTTGGGGGATGGGTGGCTGC	5640
QY	5795	CCAACTCGCTCTCCGACGCTGGTGACGTTTCGTGGGCGCGGATCCCGCAGAGCGC	5854
DB	5641	CCAACTCGCTCTCCGACGCTGGTGACGTTTCGTGGGCGCGGATCCCGCAGAGCGC	5700
QY	5855	TGTTGGGACATAGCCCTTGGGAAGTGCTCGTGGACATCTTGGCGGGCTATCGGGCGGG	5914
DB	5701	TGTTGGGACATAGCCCTTGGGAAGTGCTCGTGGACATCTTGGCGGGCTATCGGGCGGG	5760
QY	5915	GGTAGCGGGCACTCGTGGGCTTTAAGGTCAATGAGCGGAGTGCCCTCCACCGAGGA	5974
DB	5761	GGTAGCGGGCACTCGTGGGCTTTAAGGTCAATGAGCGGAGTGCCCTCCACCGAGGA	5820
QY	5975	CCTGGTCAACTTACTCCCTGCCATCTCTCTCTGGTGCCCTGCTGTGGGTCTGTGTG	6034
DB	5821	CCTGGTCAACTTACTCCCTGCCATCTCTCTCTGGTGCCCTGCTGTGGGTCTGTGTG	5880
QY	6035	CGCAGCAATATCGCTCCGCAAGTGGCGCCCGGAGAGGGGCTGTGCAGTGGATGAACCG	6094
DB	5881	CGCAGCAATATCGCTCCGCAAGTGGCGCCCGGAGAGGGGCTGTGCAGTGGATGAACCG	5940
QY	6095	GCTGATAGCGTTCGCTTCGCGGGGTAACACGCTCTCCCCACGCACTATCTGCTCAGAG	6154
DB	5941	GCTGATAGCGTTCGCTTCGCGGGGTAACACGCTCTCCCCACGCACTATCTGCTCAGAG	6000
QY	6155	CGAGCTGCAGCAGTGTCACTCAGATCTCTTAGCTTACCATCTATGTGCTCAGTGA	6214
DB	6001	CGAGCTGCAGCAGTGTCACTCAGATCTCTTAGCTTACCATCTATGTGCTCAGTGA	6060
QY	6215	GGGCTCCACAGTGGATTAATGAGGACTGCTTAGCCATGCTCCGGTCTCGTGGCTAG	6274
DB	6061	GGGCTCCACAGTGGATTAATGAGGACTGCTTAGCCATGCTCCGGTCTCGTGGCTAG	6120
QY	6275	GGATGTTTGGGATGGCAATGCACGGTGTGACTGACTTCAAGACCTGGCTCCAGTCAA	6334
DB	6121	GGATGTTTGGGATGGCAATGCACGGTGTGACTGACTTCAAGACCTGGCTCCAGTCAA	6180
QY	6335	ACTCTCCCGGGTTACCGGAGTCCCTTTCCTGTATGCCAAACGGGGTACAAGGAGT	6394
DB	6181	GCTCTCCCGGGTTACCGGAGTCCCTTTCCTGTATGCCAAACGGGGTACAAGGAGT	6240
QY	6395	CTGGCGGGGGAAGGATCATGCAAAACCACTGCCCATCGGAGACACATCGCGGACA	6454
DB	6241	CTGGCGGGGGAAGGATCATGCAAAACCACTGCCCATCGGAGACACATCGCGGACA	6300

QY	6455	TGTCAAAAACGGTTCCTCATGTAGAGATCGTAGGGCCCTAGAACTGTGAGCAACACGCTGGCACGG	6514
DB	6301	TGTCAAAAACGGTTCCTCATGTAGAGATCGTAGGGCCCTAGCAACACGCTGGCACGG	6360
QY	6515	AACGTTCCCCATCAAGCGATACACACGGGACCTTTGACACCCCTCCCGGGGCCCAACTA	6574
DB	6361	AACGTTCCCCATCAAGCGATACACACGGGACCTTTGACACCCCTCCCGGGGCCCAACTA	6420
QY	6575	TTCCAGGGCGCTATGGCGGGTGGCTGTAGGAGTACGTGAGGTTACGGCTGTGGGGG	6634
DB	6421	TTCCAGGGCGCTATGGCGGGTGGCTGTAGGAGTACGTGAGGTTACGGCTGTGGGGG	6480
QY	6635	TTTCCACTACGTGAGCGGCATGACCATGACACAGTAAAGTGCCCATGCCAGTTCCGGC	6694
DB	6481	TTTCCACTACGTGAGCGGCATGACCATGACACAGTAAAGTGCCCATGCCAGTTCCGGC	6540
QY	6695	CCCCGAATCTTACCGGAGTGGATGAGTGGCGTTGCCACAGTACGCTCCGGCTGCGAA	6754
DB	6541	CCCCGAATCTTACCGGAGTGGATGAGTGGCGTTGCCACAGTACGCTCCGGCTGCGAA	6600
QY	6755	ACCTCTTCTACGGGAGACGTCAGTTTCCAGGTGCGGGCTCAACAATATCTTGCTCGGGTC	6814
DB	6601	ACCTCTTCTACGGGAGACGTCAGTTTCCAGGTGCGGGCTCAACAATATCTTGCTCGGGTC	6660
QY	6815	CGAGTCCCATCCGAGCCGGAACGGGACGTAAACAGTCTTACTTCCATGCTACCGATCC	6874
DB	6661	CGAGTCCCATCCGAGCCGGAACGGGACGTAAACAGTCTTACTTCCATGCTACCGATCC	6720
QY	6875	CTCCCAATATACGACGAGACGCTAGCGTAGGCTGGCTAGAGGTCCTCCGCCCTCTTT	6934
DB	6721	CTCCCAATATACGACGAGACGCTAGCGTAGGCTGGCTAGAGGTCCTCCGCCCTCTTT	6780
QY	6935	AGCCAGCTCATCAGCTACGCCAGTTGTCTCGGCCCTCTTTGAAGCGACATGCATACCCA	6994
DB	6781	AGCCAGCTCATCAGCTACGCCAGTTGTCTCGGCCCTCTTTGAAGCGACATGCATACCCA	6840
QY	6995	CCATGACTCCCGGAGCGCTGACCTCATTCGAGGCGCAACCTCTTGTGGCGGACGAGATGG	7054
DB	6841	CCATGACTCCCGGAGCGCTGACCTCATTCGAGGCGCAACCTCTTGTGGCGGACGAGATGG	6900
QY	7055	CGGAAAATACATCAGTCGGTGGAGTCAGAGATAGGTAGTAAATCTTGGACTCTTTGGAACC	7114
DB	6901	CGGAAAATACATCAGTCGGTGGAGTCAGAGATAGGTAGTAAATCTTGGACTCTTTGGAACC	6960
QY	7115	GCITTCACGCGAGGGGGATCAGAGGGAGATATCCGTCGGCGGAGATCTCCGGAATAATC	7174
DB	6961	GCITTCACGCGAGGGGGATCAGAGGGAGATATCCGTCGGCGGAGATCTCCGGAATAATC	7020
QY	7175	CAGGAATTTCCCTCAGCGTTGCCATAATGAGGACATCCGTCGGCGGAGTCAATCTCCATGCT	7234
DB	7021	CAGGAATTTCCCTCAGCGTTGCCATAATGAGGACATCCGTCGGCGGAGTCAATCTCCATGCT	7080
QY	7235	AGAGTCTGAGGAGCCGGACTACGTTCCCTCGGTGGGTACACGATGCCCATTTGCCACC	7294
DB	7081	AGAGTCTGAGGAGCCGGACTACGTTCCCTCGGTGGGTACACGATGCCCATTTGCCACC	7140
QY	7295	TACCAAGGCTCTCTCCAAATACCACTCCACGGAAGAAGAGACGGTTGTCTTGCAGAGATC	7354
DB	7141	TACCAAGGCTCTCTCCAAATACCACTCCACGGAAGAAGAGACGGTTGTCTTGCAGAGATC	7200
QY	7355	CAATGTGTCTTTCGCTTGGGAGCTGGCCACTAAGACCTTTCGGTAGCTCCGGATCGTC	7414
DB	7201	CAATGTGTCTTTCGCTTGGGAGCTGGCCACTAAGACCTTTCGGTAGCTCCGGATCGTC	7260
QY	7415	GGCGCTGTATGGGACACGGGACCGCCCTTCTGACCTGGCCTCCGACGAGGTGACAA	7474
DB	7261	GGCGCTGTATGGGACACGGGACCGCCCTTCTGACCTGGCCTCCGACGAGGTGACAA	7320
QY	7475	AGGATCCGACGTTGAGTGCATCTCTCCATGCCCCCTTGAAGGGGAGCGGGGACC	7534
DB	7321	AGGATCCGACGTTGAGTGCATCTCTCCATGCCCCCTTGAAGGGGAGCGGGGACC	7380
QY	7535	CGAATCTACGCGCGGCTTTGGTCTTACCGTAGTGAGGAGGCTAGTGAGGATGTGCTGCTG	7594

Db 7381 CGATCTCAGCAGCGGCTTGGTCTACCGTGAGTGAGAGGCTAGTCAGGACGTCGCTG 7440
 Qy 7595 CTGCTCAATGTCCTATAGTGGACAGGCGCCCTGATCAGCGCATCGCGTGGGAGGAAG 7654
 Db 7441 CTGCTCAATGTCCTATAGTGGACAGGCGCCCTGATCAGCGCATCGCGTGGGAGGAAG 7500
 Qy 7655 TAAGCTGCGCATCAACCGCTTGAGCAACTCTTTCTGCTGCTCAACCAACATGCTACGC 7714
 Db 7501 TAAGCTGCGCATCAACCGCTTGAGCAACTCTTTCTGCTGCTCAACCAACATGCTACGC 7560
 Qy 7715 CACAACATCCGCGCGCGCAAGCTCCGCGAGAGAGGTCACTTTCACAGATTGGAAGT 7774
 Db 7561 CACAACATCCGCGCGCGCAAGCTCCGCGAGAGAGGTCACTTTCACAGATTGGAAGT 7620
 Qy 7775 CCTCGATGATCATACCGGAGCGTACTCAAGAGATGAAGCGGAGCGCTCCACAGTTAA 7834
 Db 7621 CCTGGACGACCATTAACCGGAGCGTACTCAAGAGATGAAGCGGAGCGCTCCACAGTTAA 7680
 Qy 7835 GGCTAAGCTTCTATCTATAGGAGGCTTGAAGTGAAGCGGCGGCGGCGGCGGCGGCGG 7894
 Db 7681 GGCTAAGCTTCTATCTATAGGAGGCTTGAAGTGAAGCGGCGGCGGCGGCGGCGGCGG 7740
 Qy 7895 CAATTTGGCTATGCGGCAAGGAGCGTCCGGAACCTATCCAGAGGCGCGGCTTACACAT 7954
 Db 7741 CAATTTGGCTATGCGGCAAGGAGCGTCCGGAACCTATCCAGAGGCGCGGCTTACACAT 7800
 Qy 7955 CCGCTCGTGGGAGGACTTGTGGAAGACACTGAACACCAATTGACACCAACATCAT 8014
 Db 7801 CCGCTCGTGGGAGGACTTGTGGAAGACACTGAACACCAATTGACACCAACATCAT 7860
 Qy 8015 GGCAGAAAGTGAAGTTTCTGCTCCACACAGAGAGGAGCGGCGGCGGCGGCGGCGGCGG 8074
 Db 7861 GGCAGAAAGTGAAGTTTCTGCTCCACACAGAGAGGAGCGGCGGCGGCGGCGGCGGCGG 7920
 Qy 8075 TATCGTATCCAGACCTGGAGTTCGTGTATGCGAGAGATGCGCTTTACGAGCGGT 8134
 Db 7921 TATCGTATCCAGACCTGGAGTTCGTGTATGCGAGAGATGCGCTTTACGAGCGGT 7980
 Qy 8135 CTCACCCCTTCTCAGCGGTGATGGGCTCCTCATACGGATTTCAATCACTCCCGGCGGCGG 8194
 Db 7981 CTCACCCCTTCTCAGCGGTGATGGGCTCCTCATACGGATTTCAATCACTCCCGGCGGCGG 8040
 Qy 8195 GCGGTCGAGTCTCTGCTGATCTCGAATCAAGAAATGCGCTATGGCTTCTCAT 8254
 Db 8041 GCGGTCGAGTCTCTGCTGATCTCGAATCAAGAAATGCGCTATGGCTTCTCAT 8100
 Qy 8255 TGACACCGCTGTTTGACTCAACGCTCACTGAGAGTGACATGCTGTTGAGGAGTCAAT 8314
 Db 8101 TGACACCGCTGTTTGACTCAACGCTCACTGAGAGTGACATGCTGTTGAGGAGTCAAT 8160
 Qy 8315 TTACCATGTTGACTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8374
 Db 8161 TTACCATGTTGACTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8220
 Qy 8375 GCTTTACATCGGCGGTCCTCTGACTCACTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8434
 Db 8221 GCTTTACATCGGCGGTCCTCTGACTCACTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8280
 Qy 8435 CCGCGAAGTGGCGTCTGACGACTAGCTCGGTAATACCTCACATGTTACTTGAAGCG 8494
 Db 8281 CCGCGAAGTGGCGTCTGACGACTAGCTCGGTAATACCTCACATGTTACTTGAAGCG 8340
 Qy 8495 CACTGAGCTGTCGAGCTCAAGCTCCAGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8554
 Db 8341 CACTGAGCTGTCGAGCTCAAGCTCCAGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8400
 Qy 8555 CTTTGTCTGTTATCTGTAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8614
 Db 8401 CTTTGTCTGTTATCTGTAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8460
 Qy 8615 CACGAGGCTATGACTAGTATTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8674

Db 8461 CACGAGGCTATGACTAGTATTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8520
 Qy 8675 CCTGGAGCTGATAAATCATCTGTTCTCCATGTCAGTCGCGCAGCATGTCATCTGCGAA 8734
 Db 8521 CCTGGAGCTGATAAATCATCTGTTCTCCATGTCAGTCGCGCAGCATGTCATCTGCGAA 8580
 Qy 8735 AAGGGTATACCTCACTCACCCTGACCGCCACCCCGCTTGCACGGGCTGCGTGGGAGAC 8794
 Db 8581 AAGGGTATACCTCACTCACCCTGACCGCCACCCCGCTTGCACGGGCTGCGTGGGAGAC 8640
 Qy 8795 AGCTAGACACATCCAAATCACTCTTGGCTAGGCAATATCATGATGTCGCGCCACCT 8854
 Db 8641 AGCTAGACACATCCAAATCACTCTTGGCTAGGCAATATCATGATGTCGCGCCACCT 8700
 Qy 8855 ATCGGCAAGGATGATCTCATGACTCACTTTTCTCCATCTTCTAGCTCAAGAGCACT 8914
 Db 8701 ATCGGCAAGGATGATCTCATGACTCACTTTTCTCCATCTTCTAGCTCAAGAGCACT 8760
 Qy 8915 TGAAGGCGCTGGATTGCA 8935
 Db 8761 TGAAGGCGCTGGATTGCA 8781

RESULT 14
 AF054259
 LOCUS
 DEFINITION
 Hepatitis C virus HC-J4, pcV-J4l10 polyprotein gene, partial cds.
 ACCESSION
 AF054259
 VERSION
 AF054259.1 GI:3098654
 KEYWORDS
 Hepatitis C virus
 SOURCE
 Hepatitis C virus
 ORGANISM
 Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (bases 1 to 8780)
 AUTHORS
 Yanagi,M., St Claire,M., Shapiro,M., Emerson,S.U., Purcell,R.H. and Bukh,J.
 TITLE
 Transcripts of a chimeric cDNA clone of hepatitis C virus genotype 1b are infectious in vivo
 JOURNAL
 Virology 244 (1), 161-172 (1998)
 MEDLINE
 98240944
 PUBMED
 9581788

REFERENCE 2 (bases 1 to 8780)
 AUTHORS
 Yanagi,M. and Bukh,J.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (18-MAR-1998) Hepatitis Viruses Section, LID, NIAID, NIH,
 7 Center Dr., Building 7, Room 201, Bethesda, MD 20892, USA

FEATURES
 source
 1..8780
 /organism="Hepatitis C virus"
 /mol_type="genomic RNA"
 /strain="HC-J4"
 /isolate="pcV-J4l10"
 /db_xref="taxon:11103"
 187..>8780
 /codon_start=1
 /product="polyprotein"
 /protein_id="AAC15732.1"
 /db_xref="GI:3098655"
 /translation="MSTNPKPQKTKRNTNRPQDVKPFPGGQIVGGVILPRRGRRL
 GYATRKYSRQPRKQPIPKAPQSGRAWQPGYWPVLTGNSGLGAGWLLSPRG
 SPNSGPIPRRRNSLKGVIDTLTCGFADLMGYIPLVGPALGGAALAHGVRIED
 GYVATGNIPLGCSFSIFLLALLSLTI PASAYEVHNSGIVHTVNDNSISVYAAD
 VIMTPEGVPCVPEADSSRCWALTPLAARNASVPTTIRRHVDLLVGTAFCSAMV
 VDULGSIPLNSQLFSPRHETVQDCNSIYPGVSHRMWMDMMNWSPTTALV
 SOLLRIPOAVDMVAGAHWGLAGLAYYSMTGNWAKVLIALLFTGVDGTYTSGAA
 CRTTSFGLSPSSGASQKILQVNTGNSHIALNCNDSIRHGLPDLAALFYTHKNS
 GCPERNASCRPIDWFGWGWPITYEPNSSDPYCMHYAPRPGIVPASQVQEVYC
 FTSPVVGVTDRSGPTYSWGENETVLLNTPFPQGNWFGCTWMNSTGPTKTCGG
 PFCNTGGVGNRTLI CPTDCFKHPEATYKCGSGFWLTPRLVDYPRYLWPTCTLNF
 STFKVRYVGGVEHRLNACNWRGCRNLEDRSELSPLLSTWQILPCATLNF
 PALSTGLHLHQNIVDVLGVGSGFSPALKWEYIILLFILLADARVCACLWMILL
 TAQRAALENLVVLNAAAVAGAHGILSFVTFCAWYIKGRAPGAATFYGVWPLLL

CDS

ILLALPPRAYALDRERMAASCGGAVLGVFLFLSLPYKVFLLRLIWLQYPTIRAEAH
 MQVWVPLNVRGRDAIILLTCAVHPELIFDTIKLLAILLGLPLMWAHLRIVPVFVR
 AQGLLRACMLNVRKAGHYVQWAFMKLIGALFTQYVYVNHITPLRDWAHLRDLAVAVE
 PVFSDMETKVTIWKADTAACGDIILGLPVSAARKGEFLGPAADSLEQGRHLLAPIT
 TVSQCTRGVLGCIILNLSRGRDNQVEGVQVSTATHFSFLAICINGVCTVYHAGSK
 TLAGEKGIPTQWNTNVDLGLWQAPPGARSMTPCSGSSDLYLVRHADYIPVRRRG
 DSRGSLSPRPVSLKSGSGGLCPGSHGVVFAAVCTRGVAKAVDFIPVESMETT
 MRSFVTDNSPFAVPQTFQVHLHAPTSGSKSTKVPAAAYAAQGVKVLVLPNSVAATL
 GFGYMSKAHGIDNIRTVGRTITGGSTVSTYGVFLADGGCSGCAVDIICDECHS
 TDSITLIGTIVLQDAETAGARLVLTATPSTVSTVPHENIEBISGLSNGBELPYGX
 AIPTEIKGGSHLIFCHSKKDELAALGLGLNNAVAYRGLDVSLPIPGDVVVA
 TDALMTGTGDFSDVIDCWTCTQVDFSLDPTFIETITVPQDAVSSQRGRGTGG
 BSGIYRVPSPRPSGKMFSSVLCYCDAGCAWELTPTASTVRLRYLNTPLGPVQC
 DHELFWSBFTGLTHIDAFISQTKQAGNFPYVAYQATVCARAQAQPPPSWDQWKC
 LIRUKTTHGFTPLLYRLAGVQNEVILTHPIITKYMCAWSADLEVVWTVLWGVGLA
 ALAAJCLTGTGVVIVRIILSKPAVDPREVLQEPDEMCEASQLPIEQGMOLAE
 OFKQKALGLLTQAKQAAAPVVEKWRLETPWAKHWNFIISGIQYLAGLSLTPGN
 PAIASLMAFTASITPLTQNTLLFNILGHWAAQLAPPSSAASFVAGIAGAAVGS
 GLGXVLDIILAGVAGVAGALVAFKVMGSEVFSTEDLVNLLPAILSPALVGVCAA
 ILRHVGEGEGAVQMMNRLIAPARGNEHVSPTHYPSDAAARVQILSLITLQLLK
 RLHOWINEDCSTPCSGSLRDVMDNICVLTFDKTQSLKLLPRLPGVPFLSCQGVK
 GWRGDDGIMQNTPCGQIAGHKVNGSMRIVGPRTCSNTWHTPPIINAYTGPCTPSP
 APNTSRALWRVAEYEVTRVGVFHYVVGITTDNVKPCQVPAPEPTEVDGVRLHR
 YAPACKPLREDVAFQGLNOLVGSOLPCPEPEDETVILTSMLJDPDSHITAEATREL
 ARGSPPLSASSASQLSAPSKAICTTHDSFDADLLEANLLMQEONGNITRVESN
 KVVILDFBPLHGDGDEREISVAAEILRKSRKFSALFIWAPDYNPLLESWMDPY
 VPPVHGCPUPPTKAPPIPPRRKRTVVLTESNVSSALAEIATKTFGSSGSADVSTQ
 ATALPDQASDDGKGSVSESYSPLEGEFGDPLSDGSMSTVSEASSEDVVCSSMS
 YTWGALITPCAAESKLPINPLNSLRLHNNMVTATTSBSASLRQKVTFDRLQVLD
 DHYRDLVKEMKASTVKALISTEAKCLTPHSAKSKFGYGHARDVNLSSRAVNH
 RSVWEDLELETERPIDITIMAKSEVFCQPEKGRKPARLIVFDLGVCRKXALYD
 VVSTLPOAVGSSVGFQYSPQRQVETWKKKPMFSDYTRDSTVTSRDIRV
 EESIYCCDLAPARQAIRLSLTERLYTGGPLTNSKQNGYRRCRAGSVLITSGNTL
 TCYLKATRAAKLOQCTMLVNGDLVITCESAGTOEDAAALRAPAEAMTRYSAPEG
 DPQPEYDLELITSCSNVSAHDAQKRVYILTRDPTPLARAAMTARHTPINSHL
 GNIIMYAPTLLWARMLMTHFILLSAQEQLEKALDC"

ORIGIN

Query Match 89.6%; Score 8599.2; DB 14; Length 8780;
 Best Local Similarity 98.7%; Pred. No. 0;
 Matches 8667; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY	156	CCGGTGAGTACACCGAATTCACAGACGACCGGCTCTTTTGGATCAACCGCTCAA	215
Db	1	CCGGTGAGTACACCGAATTCACAGACGACCGGCTCTTTTGGATCAACCGCTCAA	60
QY	216	TGCTGTGAGATTTGGGCGTCCCGCGAGACTGCTAGCCGAGTAGTGTGGGTGCGAA	275
Db	61	TGCTGTGAGATTTGGGCGTCCCGCGAGACTGCTAGCCGAGTAGTGTGGGTGCGAA	120
QY	276	AGSCCTGTGCTACTGCTATAGGTGGTTCGAGTGTCCCGGAGGCTCTCGTAGACCG	335
Db	121	AGSCCTGTGCTACTGCTATAGGTGGTTCGAGTGTCCCGGAGGCTCTCGTAGACCG	180
QY	336	TGCACCATGAGCAGCACTTAAACCTCAAGAAAACCAAGCTTAACCAACCGCGC	395
Db	181	TGCACCATGAGCAGCACTTAAACCTCAAGAAAACCAAGCTTAACCAACCGCGC	240
QY	396	CCACAGGAGTCAAGTTCCGGGCGGTGTGTCAGATCGTTGCTGAGATTACCTGTGGCG	455
Db	241	CCACAGGAGTCAAGTTCCGGGCGGTGTGTCAGATCGTTGCTGAGATTACCTGTGGCG	300
QY	456	CGCAGGGGCGCGAGTGTGGTGTGCGCGGACTAGGAGCTTCCGAGCGGTGCGAACCT	515
Db	301	CGCAGGGGCGCGAGTGTGGTGTGCGCGGACTAGGAGCTTCCGAGCGGTGCGAACCT	360
QY	516	CETGGAAGCGCACTATCCCAAGGCTTCGCCGACCGAGGCGAGGCGCTGGGTCTAG	575
Db	361	CETGGAAGCGCACTATCCCAAGGCTTCGCCGACCGAGGCGAGGCGCTGGGTCTAG	420
QY	576	CCGGGTACCCCTTGGCCCTCTATGGCAATGAGGGCTTGGGTGGGAGGATGGCTCCG	635
Db	421	CCGGGTACCCCTTGGCCCTCTATGGCAATGAGGGCTTGGGTGGGAGGATGGCTCCG	480

QY	636	TCACCCCGCGCTCCGGCTCTAGTTGGGCCCCACGACCCCCCGCGTAGTTCGGGTAAAC	695
Db	481	TCACCCCGCGCTCCGGCTCTAGTTGGGCCCCACGACCCCCCGCGTAGTTCGGGTAAAC	540
QY	696	TTGGTGAAGTCACTCGATACCTTTACATCGGCTTCGCGATCTCATGGGTACATTCCG	755
Db	541	TTGGTGAAGTCACTCGATACCTTTACATCGGCTTCGCGATCTCATGGGTACATTCCG	600
QY	756	CTCGTCGCGCCCCCTTAGGGGCGCTGCCAGGGCTTTGGCACACGGTTCGGGTCTCTG	815
Db	601	CTCGTCGCGCCCCCTTAGGGGCGCTGCCAGGGCTTTGGCACACGGTTCGGGTCTCTG	660
QY	816	GAGACCGCTGAACATGCAACAGGAACTTGCCCGGTTCGCTCTCTATCTATCTTCCCT	875
Db	661	GAGACCGCTGAACATGCAACAGGAACTTGCCCGGTTCGCTCTCTATCTATCTTCCCT	720
QY	876	TTGGCTCTGCTGTCTCTGTTGACCATCCAGCTTCCTGCTTATGAAGTGGCAACGCTCC	935
Db	721	TTGGCTCTGCTGTCTCTGTTGACCATCCAGCTTCCTGCTTATGAAGTGGCAACGCTCC	780
QY	936	GGGATATACCATGTACGAAACGACTGCTCCAACTCAAGCATTTGTATGAGGAGCGGAC	995
Db	781	GGGATATACCATGTACGAAACGACTGCTCCAACTCAAGCATTTGTATGAGGAGCGGAC	840
QY	996	GTGATCATGCATCTCCCGGTGGTGGCTGCTGTTCAGAGGGTAAACGCTCCCGTTC	1055
Db	841	GTGATCATGCATCTCCCGGTGGTGGCTGCTGTTCAGAGGGTAAACGCTCCCGTTC	900
QY	1056	TGGGTAGCGCTACTCCACGCTCGGGGCGAGAAATGCCAGCGTCCCACTACGACAATA	1115
Db	901	TGGGTAGCGCTACTCCACGCTCGGGGCGAGAAATGCCAGCGTCCCACTACGACAATA	960
QY	1116	CGAGCGCACTGCTGCTGCTGTGGGACGCTGCTTTCGCTCCGCTATGTAAGTGGGG	1175
Db	961	CGAGCGCACTGCTGCTGCTGTGGGACGCTGCTTTCGCTCCGCTATGTAAGTGGGG	1020
QY	1176	GATCTCTGGGATCTATTTCTCTGCTCCAGCTGCTTCACTTCTCGCTCCCGGCTAT	1235
Db	1021	GATCTCTGGGATCTATTTCTCTGCTCCAGCTGCTTCACTTCTCGCTCCCGGCTAT	1080
QY	1236	GAGACAGTGCAGACTGCAACTGCTCAATCTATCCCGGCGCATGATATCAGGTACCCGCTG	1295
Db	1081	GAGACAGTGCAGACTGCAACTGCTCAATCTATCCCGGCGCATGATATCAGGTACCCGCTG	1140
QY	1296	GCTTGGGATATGATGATGAACTGCTCACTCAACAGCCCTAGTGTGTGCGAGTTGCTC	1355
Db	1141	GCTTGGGATATGATGATGAACTGCTCACTCAACAGCCCTAGTGTGTGCGAGTTGCTC	1200
QY	1356	CGGATCCCAAGCTGCTGCTGGACATGTTGGGGGGCGGCTGGGAGTTCCTGGCGGG	1415
Db	1201	CGGATCCCAAGCTGCTGCTGGACATGTTGGGGGGCGGCTGGGAGTTCCTGGCGGG	1260
QY	1416	CTTGCTCTACTATTTCCATGTTAGGGAACCTGGGCTAAGGTTCTGATTTGGCGCTACTCTTT	1475
Db	1261	CTTGCTCTACTATTTCCATGTTAGGGAACCTGGGCTAAGGTTCTGATTTGGCGCTACTCTTT	1320
QY	1476	GGGGGCTTGAAGGGAGACCCACAGCGGGGAGGTGGCGGCGACACCACTCCCGGG	1535
Db	1321	ACCGGCTTGAAGGGAGACCCACAGCGGGGAGGTGGCGGCGACACCACTCCCGGG	1380
QY	1536	TTCACTCTCTCTTTCTCATCTGGGGGCTCTCAGAAAATCCAGCTTGTGATACCAACGGC	1595
Db	1381	TTCACTCTCTCTTTCTCATCTGGGGGCTCTCAGAAAATCCAGCTTGTGATACCAACGGC	1440
QY	1596	AGCTGGCAATCAACAGGACTGCCCTAAATTTGCAATGACTCCCTCCAACTGGGTCTCTT	1655
Db	1441	AGCTGGCAATCAACAGGACTGCCCTAAATTTGCAATGACTCCCTCCAACTGGGTCTCTT	1500
QY	1656	GGCGGCTGTTTACGACACAGTTTCACTGCTCCGGTGGCGGAGCGCATGGCCAGC	1715
Db	1501	GGCGGCTGTTTACGACACAGTTTCACTGCTCCGGTGGCGGAGCGCATGGCCAGC	1560

QY 1716 TGGCGCCCAATGATGTTGTCGCCAGGGGTGGGCGCCCATCACTATACCTAAAGCTAAC 1775
 Db 1561 TGGCGCCCAATGATGTTGTCGCCAGGGATGGGCGCCCATCACTATACCTAAAGCTAAC 1620
 QY 1776 AGCTCGGATCAGAGCCCTTATGCTGGCAATTAAGCGGCTCAACCGTGTGGTGTGCTACCC 1835
 Db 1621 AGCTCGGATCAGAGCCCTTATGCTGGCAATTAAGCGGCTCAACCGTGTGGTGTGCTACCC 1680
 QY 1836 GCGTCGAGGTGTGTGTCAGTGTATGTTTCACCCCAAGCCCTGTTGTGGTGGGACC 1895
 Db 1681 GCGTCGAGGTGTGTGTCAGTGTATGTTTCACCCCAAGCCCTGTTGTGGTGGGACC 1740
 QY 1896 ACCGATGTTCCGGTGTCCCTACGTATAGCTGTGGGGGAGATAGACAGAGCTGATGCTC 1955
 Db 1741 ACCGATGTTCCGGTGTCCCTACGTATAGCTGTGGGGGAGATAGACAGAGCTGATGCTC 1800
 QY 1956 CTCACAAACAGCGTCCGCCACAAAGGCAACTGGTTCGGCTGTACATGGATGAATAGTACT 2015
 Db 1801 CTCACAAACAGCGTCCGCCACAAAGGCAACTGGTTCGGCTGTACATGGATGAATAGTACT 1860
 QY 2016 GGGTTCATAAGAGCTGCGGAGGTCCCGGTGTAAACATCGGGGGGTCCGGTAAACCGCAC 2075
 Db 1861 GGGTTCATAAGAGCTGCGGAGGCCCCCGGTGTAAACATCGGGGGGTCCGGTAAACCGCAC 1920
 QY 2076 TTGATCTGCCCCCAGGACTGCTCCGGAAGCACCCCGAGGCTACTTACAAATGTGGC 2135
 Db 1921 TTGATCTGCCCCCAGGACTGCTCCGGAAGCACCCCGAGGCTACTTACAAATGTGGC 1980
 QY 2136 TCGGGGCCCTGGTTGACACCTAGGTGCTAGTACACTACCCATACAGCTTTGSCATAC 2195
 Db 1981 TCGGGGCCCTGGTTGACACCTAGGTGCTAGTACACTACCCATACAGCTTTGSCATAC 2040
 QY 2196 CCGTGCACCTCAATTTTTCATCTTTAAGGTAGGATGTATGTGGGGGCGTGGAGCAC 2255
 Db 2041 CCGTGCACCTCAATTTTTCATCTTTAAGGTAGGATGTATGTGGGGGCGTGGAGCAC 2100
 QY 2256 AGGCTCAATGCGCAGCAATGGACTCGAGAGAGCGCTGTAACTTGGAGGACAGGGAC 2315
 Db 2101 AGGCTCAATGCGCAGCAATGGACTCGAGAGAGCGCTGTAACTTGGAGGACAGGGAC 2160
 QY 2316 AGGTCAGAACTACGCGCTGTCTGTACAAACAGAGTGGCAGATACCTGCTGTGCT 2375
 Db 2161 AGGTCAGAACTACGCGCTGTCTGTACAAACAGAGTGGCAGATACCTGCTGTGCT 2220
 QY 2376 TTACACACCTACCGGCTTTATCCATGTTTATCCATGTTTATCCATCAGAAATCGTGGAC 2435
 Db 2221 TTACACACCTACCGGCTTTATCCATGTTTATCCATGTTTATCCATCAGAAATCGTGGAC 2280
 QY 2436 GTGCAATACCTGTACGGTGTAGGCTCAGCGTTGCTGTACAAACAGAGTGGCAGATAC 2495
 Db 2281 GTGCAATACCTGTACGGTGTAGGCTCAGCGTTGCTGTACAAACAGAGTGGCAGATAC 2340
 QY 2496 ATCCTGTGCTTTTCTCTCTGGCAGACGCGGCTGTGTGCTGTGCTGTGCTGTGCT 2555
 Db 2341 ATCCTGTGCTTTTCTCTCTGGCAGACGCGGCTGTGTGCTGTGCTGTGCTGTGCT 2400
 QY 2556 CTGCTGTACCGAGCTAGCGGCTGTAGAACTTGGTGTCTCTCAATGGCGGTCC 2615
 Db 2401 CTGCTGTACCGAGCTAGCGGCTGTAGAACTTGGTGTCTCTCAATGGCGGTCC 2460
 QY 2616 GTGGCGGAGCGCATGTATCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2675
 Db 2461 GTGGCGGAGCGCATGTATCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2520
 QY 2676 AAGGGAGGCTGCTCTCTGGGGCGGCTATGCTTTTATGGCGTATGGCGCTGCTCCCTG 2735
 Db 2521 AAGGGAGGCTGCTCTCTGGGGCGGCTATGCTTTTATGGCGTATGGCGCTGCTCCCTG 2580
 QY 2736 CTCTACTGCGCTTACACACAGAGCTTACGCTTGGACCGGGAGATGGCTGACAGTGC 2795
 Db 2581 CTCTACTGCGCTTACACACAGAGCTTACGCTTGGACCGGGAGATGGCTGACAGTGC 2640
 QY 2796 GGGGGTGGGCTTCTGTAGTCTGGTATCTTGTGACCTTGTACCATACTACAAAGTGT 2855

Db 2641 GGGGGTGGGCTTCTGTAGTCTGGTATCTTGTGACCTTGTACCATACTAAAGTGT 2700
 QY 2856 CTCACTAGGCTCATATGTTGTTTACAATATCTTATCACAGAGCGGAGGCGCATCAAA 2915
 Db 2701 CTCACTAGGCTCATATGTTGTTTACAATATCTTATCACAGAGCGGAGGCGCATCAAA 2760
 QY 2916 GTGTGGTCCCCCCTCAACGTTTGGGAGGCGGATGCGCATCATCTCTCTCAAGTGT 2975
 Db 2761 GTGTGGTCCCCCCTCAACGTTTGGGAGGCGGATGCGCATCATCTCTCTCAAGTGT 2820
 QY 2976 GCGGTTTCATCCAGAGTTAATTTTTCATCACCAAACTCTCTCTCGCCATCATCTCGCCCG 3035
 Db 2821 GCGGTTTCATCCAGAGTTAATTTTTCATCACCAAACTCTCTCTCGCCATCATCTCGCCCG 2880
 QY 3036 CTATGTGTCTCCAGGCTGGCATTAACAGAGTGCCTTATCTGTGTGCGCTCAAGGGCTC 3095
 Db 2881 CTATGTGTCTCCAGGCTGGCATTAACAGAGTGCCTTATCTGTGTGCGCTCAAGGGCTC 2940
 QY 3096 ATTCTGTGATGATGTGTGCGAAAGTGCAGGGGCTCATTTATGTCANATGTCTTC 3155
 Db 2941 ATTCTGTGATGATGTGTGCGAAAGTGCAGGGGCTCATTTATGTCANATGTCTTC 3000
 QY 3156 ATGAAGCTGGGCGCTGACAGGTACGTACGTTTATAACCATCTTACCCCACTGCGGGAC 3215
 Db 3001 ATGAAGCTGGGCGCTGACAGGTACGTACGTTTATAACCATCTTACCCCACTGCGGGAC 3060
 QY 3216 TGGGCCCAACCGGCTTACAGACCTTGTGGTGGCGGTAGAGCCGCTGTCTTCGCGC 3275
 Db 3061 TGGGCCCAACCGGCTTACAGACCTTGTGGTGGCGGTAGAGCCGCTGTCTTCGCGC 3120
 QY 3276 ATGAGACCAAGGTCACTACCTGCGGAGCAGACACCGCTGCTGTGGGACATCATCTTCG 3335
 Db 3121 ATGAGACCAAGGTCACTACCTGCGGAGCAGACACCGCTGCTGTGGGACATCATCTTCG 3180
 QY 3336 GGTCTACCTCTCTCCCGGAGGAGGAGATATTTTGGGACCGGCTGATAGTCTC 3395
 Db 3181 GGTCTACCTCTCTCCCGGAGGAGGAGATATTTTGGGACCGGCTGATAGTCTC 3240
 QY 3396 GAAGGCAAGGTTGGGCTCTCTTGGGCCATCAGGCTCTCTCCCAACAGCGGGGCG 3455
 Db 3241 GAAGGCAAGGTTGGGCTCTCTTGGGCCATCAGGCTCTCTCCCAACAGCGGGGCG 3300
 QY 3456 GTACTTGTGTGATCATCATCTAGCTTACAGCGCGGAGCAAGAACAGGTTCGAAGGGAG 3515
 Db 3301 GTACTTGTGTGATCATCATCTAGCTTACAGCGCGGAGCAAGAACAGGTTCGAAGGGAG 3360
 QY 3516 GTTCAAGTGGTTTCTACCGCAACAAATCTTCTTGGGACCTGATCAACGGCGTGTGC 3575
 Db 3361 GTTCAAGTGGTTTCTACCGCAACAAATCTTCTTGGGACCTGATCAACGGCGTGTGC 3420
 QY 3576 TGGACTGTCTACCATGGCGCTGGTTCGAGACCTCTGCTGTGGCGGCTCAAAAGTTCATAC 3635
 Db 3421 TGGACTGTCTACCATGGCGCTGGTTCGAGACCTCTGCTGTGGCGGCTCAAAAGTTCATAC 3480
 QY 3636 CAATGTACACCAATGTAGACCTCTGCGGCTGGCAGCGGCGGCGGCGGCGGCG 3695
 Db 3481 CAATGTACACCAATGTAGACCTCTGCGGCTGGCAGCGGCGGCGGCGGCGGCGGCG 3540
 QY 3696 TCCATGACACCAATGTAGAGCTGTGGAGCTTACTTGGTCAAGAGACATGTCTGAT 3755
 Db 3541 TCCATGACACCAATGTAGAGCTGTGGAGCTTACTTGGTCAAGAGACATGTCTGAT 3600
 QY 3756 GTCAATCGGTGCGCGGCGGAGGAGCAGAGGGAGTCTACTCTCCCGAGGCGGCTC 3815
 Db 3601 GTCAATCGGTGCGCGGCGGAGGAGCAGAGGGAGTCTACTCTCCCGAGGCGGCTC 3660
 QY 3816 TCCCTACTGAAAGGCTCTCTCGGTTGCTTCCCTTGGGAGCAGCTGTGTGGG 3875
 Db 3661 TCCCTACTGAAAGGCTCTCTCGGTTGCTTCCCTTGGGAGCAGCTGTGTGGG 3720
 QY 3876 GTCTTCCGGGCTGTGTGTGACCGCGGGGTGCGAAGCGGTGGACTTCAATACCGGTT 3935

QY 6096 CTGATAGCGTTTCGCTTCGCGGGTAACCAACGCTCCCTACGCACTATGTCGCTGAGGC 6155
Db 5941 CTGATAGCGTTTCGCTTCGCGGGTAACCAACGCTCCCTACGCACTATGTCGCTGAGGC 6000
QY 6156 GACCGTGCAGCAGTGTCACATCAGATCCTCTCTAGCCTTACCATCACTCAACTCTGAAG 6215
Db 6001 GACCGTGCAGCAGTGTCACATCAGATCCTCTCTAGCCTTACCATCACTCAACTCTGAAG 6060
QY 6216 CCGCTCCACCAAGTGGATTAATGAGGACTGCTCTAGGCACTCTCGGCTGTGGCTTAAG 6275
Db 6061 CCGCTTCACCAATGATCAATGAGGACTGCTCTAGGCACTCTCGGCTGTGGCTTAAG 6120
QY 6276 GATGTTTGGGATGGATATGACACGGTGTGACTGACTTCAAGACTTGCCTCCAGTCCAAA 6335
Db 6121 GATGTTTGGGATGGATATGACACGGTGTGACTGACTTCAAGACTTGCCTCCAGTCCAAA 6180
QY 6336 CTCCTGCGCGGTTACCGGGAGTCCCTTCTCTGTCATGCGCNAACCGGGTACAGGGATC 6395
Db 6181 CTCCTGCGCGGTTACCGGGAGTCCCTTCTCTGTCATGCGCAACCGGGTACAGGGATC 6240
QY 6396 TGGCGGGGACGGCATCATCAAAACACCTGCCCATGCGGAGCACAGATCGCCGGACAT 6455
Db 6241 TGGCGGGGAGTGCATCATGCAAAACACCTGCCCATGCGGAGCACAGATCGCCGGACAT 6300
QY 6456 GTCAAAAACGGTTCCATGAGGATCGTAGGGCTTAGAACCTTGACAAACAGTGGCACGGA 6515
Db 6301 GTCAAAAACGGTTCCATGAGGATCGTAGGGCTTAGAACCTTGACAAACAGTGGCACGGA 6360
QY 6516 ACGTTCCCATCAACGATACACACGGGACTTGCACACCTCCCGGGGCCCACTAT 6575
Db 6361 ACATTCCTCCCATCAACGATACACACGGGACTTGCACACCTCCCGGGGCCCACTAT 6420
QY 6576 TCCAGGCGGTTATGGCGGGTGGCTGTGAGGATGACGTGAGGTTACGCTGTGGGGAT 6635
Db 6421 TCCAGGCGGTTATGGCGGGTGGCTGTGAGGATGACGTGAGGTTACGCTGTGGGGAT 6480
QY 6636 TTCCACTAGTGCAGGGCATGACCACTGACAACTGAAAGTGCACATGCCAGGTTCCGGCC 6695
Db 6481 TTCCACTAGTGCAGGGCATGACCACTGACAACTGAAAGTGCACATGCCAGGTTCCGGCC 6540
QY 6696 CCGCAATCTTACGGAGGTGGATGGATGGGTTGACAGGTACGTCACAGGTACGCTCGGGTGCAAA 6755
Db 6541 CCGCAATCTTACGGAGGTGGATGGATGGGTTGACAGGTACGTCACAGGTACGCTCGGGTGCAAA 6600
QY 6756 CCTCTTCTACGGAGGACGTCACCTTCCAGGTGGGCTCAACCAATACTTGGTGGGTGCG 6815
Db 6601 CCGCTCTCTACGGAGGACGTCGCTTCCAGGTGGGCTCAACCAATACTTGGTGGGTGCG 6660
QY 6816 CAGCTCCCATGCGAGCCCGAACCGGACGTAACAGTGTCTTCTCCATGCTCAACGATPCC 6875
Db 6661 CAGCTCCCATGCGAGCCCGAACCGGACGTAACAGTGTCTTCTCCATGCTCAACGATPCC 6720
QY 6876 TCCACATTAACAGAGAGCGGTAAAGGTGGCTAGAGGTCTCCCGCTCTTTA 6935
Db 6721 TCCACATTAACAGAGAGCGGTAAAGGTGGCTAGAGGTCTCCCGCTCTTTA 6780
QY 6936 GCGAGCTCATAGTACCGAAGTGTCTGCGCTTCTTTGAAGGCGCATGCACTACCCAC 6995
Db 6781 GCGAGCTCATAGTACCGAAGTGTCTGCGCTTCTTTGAAGGCGCATGCACTACCCAC 6840
QY 6996 CATGACTCCCGGACGCTGACCTCATCGAGGCAACCTCTTGGCGGAGAGATGGGC 7055
Db 6841 CATGACTCCCGGACGCTGACCTCATCGAGGCAACCTCTTGGCGGAGAGATGGGC 6900
QY 7056 GGAACATCACTCGCTGGAGTCAGAGATAAGTGTAAATCTTGGACTCTTTCGACCG 7115
Db 6901 GGAACATCACTCGCTGGAGTCAGAGATAAGTGTAAATCTTGGACTCTTTCGACCG 6960
QY 7116 CTTCAACCGGAGGGGATGAGGGAGATATCCGTGCGCGGAGATCTTTCGCAAAATCC 7175
Db 6961 CTTCAACCGGATGGGATGAGGGAGATATCCGTGCGCGGAGATCTTTCGCAAAATCC 7020
QY 7176 ASGAAGTTCCTCCCTCAGCGTTGCCCATATGCGGACGCCCGGACTACATCTCCACTGCTA 7235

Db 7021 AGGAGTTCCCTCAGGTTGCCCATATGGGACGCCCGGACTACATCTCCACTGCTA 7080
QY 7236 GAGTCTTGAAGAACCGGACTACGTCCTCCGCTGGTACACGGATGCCAATTCGCCACCT 7295
Db 7081 GAGTCTTGAAGAACCGGACTACGTCCTCCGCTGGTACACGGATGCCAATTCGCCACCT 7140
QY 7296 ACCAAGGCTCTCCATATACACCTCCACGGAGAAAGAGACGGTTCCTCTGACAGATCC 7355
Db 7141 ACCAAGGCTCTCCATATACACCTCCACGGAGAAAGAGACGGTTCCTCTGACAGATCC 7200
QY 7356 AATGTGCTTCTGCTTGGCGGAGCTCGCACTTAACACCTTCGTAAGTCTCCGATCGTCG 7415
Db 7201 AATGTGCTTCTGCTTGGCGGAGCTCGCACTTAAGACCTTCGTAAGTCTCCGATCGTCG 7260
QY 7416 GCGTGTATGAGGAGCGGACCGGCTCTCTGACCTGCGCTTCGACGACGAGTGAACA 7475
Db 7261 GCGTGTATGAGGAGCGGACCGGCTCTCTGACGAGCTTCGACGAGTGAACA 7320
QY 7476 GATCCGAGCTTGAGTCTGACTCTCTCATGCGCCCTTCGTAAGGAGGAGCGGGGACCC 7535
Db 7321 GATCCGAGCTTGAGTCTGACTCTCTCATGCGCCCTTCGTAAGGAGGAGCGGGGACCC 7380
QY 7536 GATCTCAGGACGGGCTTGTGCTTACCGTGAAGTGAAGGAGTGTGCTGCTG 7595
Db 7381 GATCTCAGGACGGGCTTGTGCTTACCGTGAAGTGAAGGAGTGTGCTGCTG 7440
QY 7596 TGCTCAATGCTTATACGTGAGACGCGGCTGATCAGGCAATGCGCTCGGAGGAAAGT 7655
Db 7441 TGCTCAATGCTTATACGTGAGACGCGGCTGATCAGGCAATGCGCTCGGAGGAAAGT 7500
QY 7656 AAGTCCGCACTCAACCGGTTGAGCACTCTTCTGCTGGTCAACCAACTGCTTACGCT 7715
Db 7501 AAGTCCGCACTCAACCGGTTGAGCACTCTTCTGCTGGTCAACCAACTGCTTACGCT 7560
QY 7716 ACAACATCCCGAGCGCAAGCCTCCCGGAGAGAAAGTCACTTTGACAGATTGCAAGTC 7775
Db 7561 ACAACATCCCGAGCGCAAGCCTCCCGGAGAGAAAGTCACTTTGACAGATTGCAAGTC 7620
QY 7776 CTGAGTATCATTTACCGGAGCTGCTCAAGGAGATGAAGGCGAAGCGTCCACAGTTAAG 7835
Db 7621 CTGAGTATCATTTACCGGAGCTGCTCAAGGAGATGAAGGCGAAGCGTCCACAGTTAAG 7680
QY 7836 GCTAAGCTCTATCTATAGGAGGCTGCAAGCTGACGCCCCCATTTCGCGCAATCC 7895
Db 7681 GCTAAGCTCTATCTATAGGAGGCTGCAAGCTGACGCCCCCATTTCGCGCAATCC 7740
QY 7896 AAATTTGGCTATGGGCAAAAGACGTCGGAACCTATCCAGAGGCGCGTTAAACCAATC 7955
Db 7741 AAATTTGGCTATGGGCAAAAGACGTCGGAACCTATCCAGAGGCGCGTTAAACCAATC 7800
QY 7956 CGCTCCGCTGGAGACCTTCTGAGAGACATGAAACACCAATTCACACCACTATCAG 8015
Db 7801 CGCTCCGCTGGAGACCTTCTGAGAGACATGAAACACCAATTCACACCACTATCAG 7860
QY 8016 GCAAAAGTGAGTCTTCTGCGTCCAAACAGAGAGAGGCGGCAAGCAGTGCCTT 8075
Db 7861 GCAAAAGTGAGTCTTCTGCGTCCAAACAGAGAGAGGCGGCAAGCAGTGCCTT 7920
QY 8076 ATCGTATTCAGACCTGCGAGTTCGTGATGCGAGAGATGGCTTTTACAGCTGCTC 8135
Db 7921 ATCGTATTCAGACCTGCGAGTTCGTGATGCGAGAGATGGCTTTTACAGCTGCTC 7980
QY 8136 TCCACCTTCTCAGCGCTGATGGCTCTCATACGATTTCAATCTCCCGCAAGAG 8195
Db 7981 TCCACCTTCTCAGCGCTGATGGCTCTCATACGATTTCAATCTCCCGCAAGAG 8040
QY 8196 CCGGTGAGTCTCTGCTGAATACCTGGAATCAAGAAATGCGCTTATGCGCTTCTCATAT 8255
Db 8041 CCGGTGAGTCTCTGCTGAATACCTGGAATCAAGAAATGCGCTTATGCGCTTCTCATAT 8100
QY 8256 GACACCGCTGTTTGAATCAACCGTCACTGAGTGAATTCGTGTGAGAGTCAAT 8315

QY 276 AGGCTTGTGGTACTCCCTGATAGGCTGCTGCGAGTGCCTCGGAGGCTCGTAGACCG 335
 Db 121 AGGCTTGTGGTACTCCCTGATAGGCTGCTGCGAGTGCCTCGGAGGCTCGTAGACCG 180
 QY 336 TGCACCATGAGCAGCAATCCTAAACCTCAAGAAAAACCAAAAGTAAACCAACCGCCGC 395
 Db 181 TGCACCATGAGCAGCAATCCTAAACCTCAAGAAAAACCAAAAGTAAACCAACCGCCGC 240
 QY 396 CCACAGAGCGTCAAGTTCCTGCGGCGGTGTCAGATCGTTGGTGGAGTTACCTGTTGCCG 455
 Db 241 CCACAGAGCGTCAAGTTCCTGCGGCGGTGTCAGATCGTTGGTGGAGTTACCTGTTGCCG 300
 QY 456 CGCAGGGGCCCCAGGTGGGTGTCGCGGACCTAGGAAGCTTCCGAGCGGTGCGCAACCT 515
 Db 301 CGCAGGGGCCCCAGGTGGGTGTCGCGGCGACTAGGAAGCTTCCGAGCGGTGCGCAACCT 360
 QY 516 CPTGGAAGCGCAACCTATPCCAAAGGCTGCGGACCCGAGGCGAGGGCCCTGCGCTCAG 575
 Db 361 CPTGGAAGCGCAACCTATPCCAAAGGCTGCGGACCCGAGGCGAGGGCCCTGCGCTCAG 420
 QY 576 CCGGGTACCCCTGCGGCTTATGCGAATGAGGGCTTGGGTGCGGACGATGGCTCCTG 635
 Db 421 CCGGGTACCCCTGCGGCTTATGCGAATGAGGGCTTGGGTGCGGACGATGGCTCCTG 480
 QY 636 TCACCCCGGGCTCCCGGCTAGTTCGGGCCCCACAGGACCCCGGCGTAGGTCCGCTAAC 695
 Db 481 TCACCCCGGGCTCCCGGCTAGTTCGGGCCCCACAGGACCCCGGCGTAGGTCCGCTAAC 540
 QY 696 TTGGTAAGGTATCGATACCCCTTAATGCGGCTTTCGCGATCTCATGCGGTACATTCG 755
 Db 541 TTGGTAAGGTATCGATACCCCTTAATGCGGCTTTCGCGATCTCATGCGGTACATTCG 600
 QY 756 CTGTCGGGCCCCCTAGGGGCGCTGCGAGGCTTGGCACACGCTGTCGGGCTCTG 815
 Db 601 CTGTCGGGCCCCCTAGGGGCGCTGCGAGGCTTGGCACACGCTGTCGGGCTCTG 660
 QY 816 GAGGACGGCTGAATATGCAACAGGGAATTCGCGGCTTCTCTCTATCTCTCCTC 875
 Db 661 GAGGACGGCTGAATATGCAACAGGGAATTCGCGGCTTCTCTCTATCTCTCCTC 720
 QY 876 TTGGCTCTGCTGCTGTTGACATCCAGCTTCGCTTATGAAGTGCACAGTGTCC 935
 Db 721 TTGGCTCTGCTGCTGTTGACATCCAGCTTCGCTTATGAAGTGCACAGTGTCC 780
 QY 936 GGCATATACATGTCAGAAAGTCTGCTCAACTCAAGCAATGCTATGAGGCGAGCGGAC 995
 Db 781 GGCATATACATGTCAGAAAGTCTGCTCAACTCAAGCAATGCTATGAGGCGAGCGGAC 840
 QY 996 GTGATCATGCATATCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
 Db 841 GTGATCATGCATATCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 QY 1056 TGGGTAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1115
 Db 901 TGGGTAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 QY 1116 CGACGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1175
 Db 961 CGACGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 QY 1176 GATCTCTGCGATATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
 Db 1021 GATCTCTGCGATATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 QY 1236 GAGACGTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1295
 Db 1081 GAGACGTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 QY 1296 GCTTGGGATATGATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
 Db 1141 GCTTGGGATATGATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

QY 1356 CGGATCCCAAGCTGTCTGTGACATGTGTGGCGGGGCCCATGTGGGAGTCTCTGTGGCGGC 1415
 Db 1201 CGGATCCCAAGCTGTCTGTGACATGTGTGGCGGGGCCCATGTGGGAGTCTCTGTGGCGGC 1260
 QY 1416 CTTGGCTACTATTCATGTAGGGAATGSGCTAAGTCTGATTTGTTGGCCCTACTCTTT 1475
 Db 1261 CTTGGCTACTATTCATGTAGGGAATGSGCTAAGTCTGATTTGTTGGCCCTACTCTTT 1320
 QY 1476 GCGCGCTGTGACGGGAGACCCACACGACGGGAGGCTGGCGGGCCACACACCTCTCCGG 1535
 Db 1321 ACCGGCTGTGACGGAGCAGACCTACACGTCGGGGGGGTGGCGGGCCACACCTCTCCGG 1380
 QY 1536 TPCAGTCCCTTTCTCATCTGGGCGCTCAGAAAAATCCAGCTTGTGAATACCAACGCG 1595
 Db 1381 TPCAGTCCCTTTCTCATCTGGGCGCTCAGAAAAATCCAGCTTGTGAATACCAACGCG 1440
 QY 1596 AGCTGGCAGTCAACAGGACTGCCCTAAATTCAGATGCTCCCTCCAACTGGGTTCTTT 1655
 Db 1441 AGCTGGCAGTCAACAGGACTGCCCTAAATTCAGATGCTCCCTCCAACTGGGTTCTTT 1500
 QY 1656 GCGCGCTGTGTTACGCAACAAAGTTCACCTGCTCCGGTGGCGGGAGCGCATGSCCAGC 1715
 Db 1501 GCGCGCTGTGTTCTACACACAAAGTTCACCTGCTCCGGTGGCGGGAGCGCATGSCCAGC 1560
 QY 1716 TGCSCGCCCATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1775
 Db 1561 TGCSCGCCCATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 QY 1776 AGCTGGATCAGAGGCTTATGCTGCGATTCACGCGCTCGACCTGCTGCTGCTGCTGCTGCT 1835
 Db 1621 AGCTGGATCAGAGGCTTATGCTGCGATTCACGCGCTCGACCTGCTGCTGCTGCTGCTGCT 1680
 QY 1836 GCGTCCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1895
 Db 1681 GCGTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 QY 1896 ACCGATCGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1955
 Db 1741 ACCGATCGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 QY 1956 CTCACACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2015
 Db 1801 CTCACACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 QY 2016 GCGTTCACCTAAGACGCTGCGAGGTCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2075
 Db 1861 GCGTTCACCTAAGACGCTGCGAGGTCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 QY 2076 TTGATCTGCCCCACGAGCTGCTTCCGGAAGCACTCCGAGGCTACTTACACAAAATGTGGC 2135
 Db 1921 TTGATCTGCCCCACGAGCTGCTTCCGGAAGCACTCCGAGGCTACTTACACAAAATGTGGC 1980
 QY 2136 TCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2195
 Db 1981 TCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
 QY 2196 CCGTGCACCTCAATTTTCCATTTTCCATTTTCCATTTTCCATTTTCCATTTTCCATTTTCCAT 2255
 Db 2041 CCGTGCACCTCAATTTTCCATTTTCCATTTTCCATTTTCCATTTTCCATTTTCCATTTTCCAT 2100
 QY 2256 AGGCTCAATGCCGATGCAATGAGCTCGAGAGAGCGCTGTAACTTGGAGGACAGGAT 2315
 Db 2101 AGGCTCAATGCCGATGCAATGAGCTCGAGAGAGCGCTGTAACTTGGAGGACAGGAT 2160
 QY 2316 AGGTCAGAACTCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2375
 Db 2161 AGGTCAGAACTCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 QY 2376 TTCACACCCCTACCGGCTTATCCACTGCTTATCCACTGCTTATCCACTGCTTATCCACTGCTTAT 2435
 Db 2221 TTCACACCCCTACCGGCTTATCCACTGCTTATCCACTGCTTATCCACTGCTTATCCACTGCTTAT 2280
 QY 2436 GTGCAATACCTGTACGCTGTAGGCTCAGGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2495

6661	Db		CAGCTTCATGCGAGCCGGAACCGAGCTAACAGTGTCTTACTTCCATGCTCAACCGACCC	6720
6676	Qy		TCCACATTTACAGCAGAGAGGGCTTAACGCTAGAGCTGGCTAGAGGGTCTCCCCCTCTTTA	6935
6721	Db		TCCACATTTACAGCAGAGAGGGCTTAACGCTAGAGCTGGCTAGAGGGTCTCCCCCTCTTTA	6780
6936	Qy		GCCAGCTCATCAGCTAGCCAGTGTCTGCGCTTCTTTGAAGCGCAGCATGCATACCCAC	6995
6781	Db		GCCAGCTCATCAGCTAGCCAGTGTCTGCGCTTCTTTGAAGCGCAGCATGCATACCCAC	6840
6996	Qy		CATGACTCCCGGAGCGGTGACTCATCGAGGCCAACCTCTTTGTGGCGGAGAGATGGGC	7055
6841	Db		CATGACTCCCGGAGCGGTGACTCATCGAGGCCAACCTCTTTGTGGCGGAGAGATGGGC	6900
7056	Qy		GGAAACATCACTCGCTGGAGTCAAGAAATAAGGTAGTAATTTCTGGACTCTTTCCGAACCG	7115
6901	Db		GGAAACATCACTCGCTGGAGTCAAGAAATAAGGTAGTAATTTCTGGACTCTTTCCGAACCG	6960
7116	Qy		CTTTCACGCGGAGGGGATGAGAGGGAGATATCCCTCGCGCGGAGATCTCTCGCAAAATCC	7175
6961	Db		TTTTCACGCGGATGGGATGAGAGGGAGATATCCCTCGCGCGGAGATCTCTCGCAAAATCC	7020
7176	Qy		AGGAAGTTCCCTCAGCGTTGCCATATGCGCAGCCCGGACTACAATCTCCACTGCTA	7235
7021	Db		AGGAAGTTCCCTCAGCGTTGCCATATGCGCAGCCCGGACTACAATCTCCACTGCTA	7080
7236	Qy		GAGTCTCGGAGGACCGGACTAGTCCCTCCGCTGGTGAACGGATGCCATGTCACACT	7295
7081	Db		GAGTCTCGGAGGACCGGACTAGTCCCTCCGCTGGTGAACGGATGCCATGTCACACT	7140
7296	Qy		ACCAAGGCTCTCCAATACCACTCCACGGAGAAAGAGACGGTTGCTCTGACAGAAATCC	7355
7141	Db		ACCAAGGCTCTCCAATACCACTCCACGGAGAAAGAGACGGTTGCTCTGACAGAAATCC	7200
7356	Qy		AATGTGTCTTGCTTGCGCGAGCTCGCCACTAAGA CCTTGGTAGCTTCGGATCGTGC	7415
7201	Db		AATGTGTCTTGCTTGCGCGAGCTCGCCACTAAGA CCTTGGTAGCTTCGGATCGTGC	7260
7416	Qy		GCCCTTGATAGCGCACGGGACCGCCCTTCTGACCTGCGCTCCGACAGCGGTGACAAA	7475
7261	Db		GCCCTTGATAGCGCACGGGACCGCCCTTCTGACCGCTCCGACAGCGGTGACAAA	7320
7476	Qy		GGATCCGACGTTGAGTCTGACTCTCCATGCCCCCTTGAAGGGGAGCGGGGACCCC	7535
7321	Db		GGATCCGACGTTGAGTCTGACTCTCCATGCCCCCTTGAAGGGGAGCGGGGACCCC	7380
7536	Qy		GATCTACGACGGGTCTTGCTTACCGTGAAGTGAAGAGGCTAGTAGGATGTGCTGTGC	7595
7381	Db		GATCTACGACGGGTCTTGCTTACCGTGAAGTGAAGAGGCTAGTAGGATGTGCTGTGC	7440
7596	Qy		TGCTCAATGCTTATAGTGACAGCGCCCTGATCAGCCATCCGCTCGGAGGAAAGT	7655
7441	Db		TGCTCAATGCTTATAGTGACAGCGCCCTGATCAGCCATCCGCTCGGAGGAAAGT	7500
7656	Qy		AAGCTGCCATCAACCCGTTGAGCAACTTTGTGCTGCTACCAACAATGTTCTAGCC	7715
7501	Db		AAGCTGCCATCAACCCGTTGAGCAACTTTGTGCTGCTACCAACAATGTTCTAGCC	7560
7716	Qy		ACACATCCCGCAGGAGAGCTCCGGCAGAAGAGTTCACCTTTGACAGTTCGAATGC	7775
7561	Db		ACACATCCCGCAGGAGAGCTCCGGCAGAAGAGTTCACCTTTGACAGTTCGAATGC	7620
7776	Qy		CTGGATGATCATACCGGAGCTACTCAAGGATGAGGCGAGGCGTCCACAGTTAAG	7835
7621	Db		CTGGATGATCATACCGGAGCTACTCAAGGATGAGGCGAGGCGTCCACAGTTAAG	7680
7836	Qy		GCTAAGCTTCTATCTATAGAGGAGGCTTCAGCTGACGCCCCCAATTCGGCCAAATCC	7895
7681	Db		GCTAAGCTTCTATCTATAGAGGAGGCTTCAGCTGACGCCCCCAATTCGGCCAAATCC	7740
7896	Qy		AAATTTGGCTATGGGGCAAGGACGTCCGGAACTTATCCAGCAGGGCGGTTAACCAATC	7955

Db	7741	AAANTTTGGCTATCGGGCAAGACGTCGCGAAACCTATCCAGACGGCCGTTAACCAATC	7800
Qy	7956	CGCTCCGTCGTGGAGCACTTGCTGGAGAACACTGAAACACCAATTGACACCAACCATCATG	8015
Db	7801	CGCTCCGTCGTGGAGCACTTGCTGGAGAACACTGAAACCAATTGACACCAACCATCATG	7850
Qy	8016	GCAAAAAGTGAGGTTTTCGCTCCAAACGAGAGAGGGAGCGCGACGACGTCGCGCTT	8075
Db	7861	GCAAAAAGTGAGGTTTTCGCTCCAAACGAGAGAGGGAGCGCGACGACGTCGCGCTT	7920
Qy	8076	ATCGTATTTCCAGACCTGGAGTTCGTGTATCGAGAAAGATGGCCCTTTTACGACGTGGTC	8135
Db	7921	ATCGTATTTCCAGACCTGGAGTTCGTGTATCGAGAAAGATGGCCCTTTTACGACGTGGTC	7980
Qy	8136	TCACACCTTCCCTCAGGCGGTATGGGCTCCTCATACGATTTCAATCTCCCCCAAGCAG	8195
Db	7981	TCACACCTTCCCTCAGGCGGTATGGGCTCCTCATACGCGTTTCAATCTCCCCCAAGCAG	8040
Qy	8196	CGGTCGAGTTCCTGGTGAATACCTCGAAATCAAGAAATGCCCTATGGGCTTCTCATAT	8255
Db	8041	CGGTCGAGTTCCTGGTGAATACCTGGAAATCAAGAAATGCCCTATGGGCTTCTCATAT	8100
Qy	8256	GACACCCGCTGTTTTGACTCAACCGGTCACTGAGAGTGACATTCGTGTGGAGAGTCAATT	8315
Db	8101	GACACCCGCTGTTTTGACTCAACCGGTCACTGAGAGTGACATTCGTGTGGAGAGTCAATT	8160
Qy	8316	TACCAATGTTGTGACTTTGGCCCCGGAGGCAGACAGGCCATAAGTCGCTCACAGAGCGG	8375
Db	8161	TACCAATGTTGTGACTTTGGCCCCGGAGGCAGACAGGCCATAAGTCGCTCACAGAGCGG	8220
Qy	8376	CTTTACATCGGGGTCCCTGACTAACTCAAAAGGGCAGAACTCGCGTTATCGCCGGTGC	8435
Db	8221	CTTTACATCGGGGTCCCTGACCACTCAAAAGGGCAGAACTCGCGTTATCGCCGGTGC	8280
Qy	8436	CGGCAAGTGGCGTGTGACGATGCTGGGTAAATACCTCACATGTTTACTTTGAAGGCG	8495
Db	8281	CGGCAAGTGGCGTGTGACGATGCTGGGTAAATACCTCACATGTTTACTTTGAAGGCG	8340
Qy	8496	ACTGCACCTTCGAGCTGCACAACTCCAGACTGTCACGATGTCGTGCAACGGAGACGAC	8555
Db	8341	ACTCGGCCCTTCGAGCTGCACAACTCCAGACTGTCACGATGTCGTGCAACGGAGACGAC	8400
Qy	8556	CTTGTCTTTATCTGTGAAGCGCGGAAACCCAGAGGATGGCGCGCCCTACGAGCCTTC	8615
Db	8401	CTTGTCTTTATCTGTGAAGCGCGGAAACCCAGAGGATGGCGCGCCCTACGAGCCTTC	8460
Qy	8616	ACGAGGCTATGACTAGGTATTCGCCCCCTCCGGGGATCGCCCCCAACAGATATGAC	8675
Db	8461	ACGAGGCTATGACTAGGTATTCGCCCCCTCCGGGGATCGCCCCCAACAGATATGAC	8520
Qy	8676	CTGAGCTGTATAACATCATGTTCTCCAAATGTGTCAGTCGCGCACGATGATCTGGCAAA	8735
Db	8521	CTGAGCTGTATAACATCATGTTCTCCAAATGTGTCAGTCGCGCACGATGATCTGGCAAA	8580
Qy	8736	AGGTTATATACCTCACCCGTGACCCACACCCCTTGGACGCGGCTGGTGGGAGACA	8795
Db	8581	AGGTTATATACCTCACCCGTGACCCACACCCCTTGGACGCGGCTGGTGGGAGACA	8640
Qy	8796	GCTAGACACTCCAACTCACTTTGCTAGGCAATATCATATGATGGCCCCACCTTA	8855
Db	8641	GCTAGACACTCCAACTCACTTTGCTAGGCAATATCATATGATGGCCCCACCTTA	8700
Qy	8856	TGGCAAGGATGATTTGATGACTCACTTTTTCCTCATCTTCTTAGCTCAAGAGCAACTT	8915
Db	8701	TGGCAAGGATGATTTGATGACTCACTTTTTCCTCATCTTCTTAGCTCAAGAGCAACTT	8760
Qy	8916	GAATAAGCCCTGGATTTGCA	8935
Db	8761	GAATAAGCTCTGGATTTGCA	8780

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2004, 14:55:56 ; Search time 20675 Seconds
(without alignments)
13858.652 Million cell updates/sec

Title: US-09-662-454-4
Perfect score: 9595
Sequence: 1 gccaccccccgtatggggc.....ggcctctgcagatcatgt 9595

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estav.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vxl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	129	1.3	488	9	AV755731
c 2	98	1.0	492	9	AV758366
c 3	77.8	0.8	914	13	BQ232672
c 4	76.6	0.5	807	12	BQ108244

c 5	76	0.8	796	29	CNS011GK
c 6	74.8	0.8	548	12	BM812285
c 7	74.4	0.8	562	13	BQ747212
c 8	74.4	0.8	979	10	BF035653
c 9	74.2	0.8	656	11	BC033047
c 10	74.2	0.8	848	14	CA758134
c 11	73.8	0.8	1003	12	BM415727
c 12	73.4	0.8	566	14	CF754016
c 13	73.4	0.8	731	12	BG579884
c 14	73	0.8	202	10	BP814420
c 15	73	0.8	558	9	AL643867
c 16	72.8	0.8	183	12	BG996024
c 17	72.8	0.8	664	10	BF144889
c 18	72.8	0.8	985	13	BX432529
c 19	72.4	0.8	407	9	AV650703
c 20	72.4	0.8	943	12	BG392408
c 21	72.2	0.8	428	13	BY031133
c 22	72	0.8	465	10	BE752126
c 23	72	0.8	473	12	BI814061
c 24	72	0.8	575	12	BG924378
c 25	72	0.8	898	13	BUS51262
c 26	72	0.8	994	13	BX370084
c 27	72	0.8	1049	14	CF4560C0
c 28	71.8	0.7	198	14	CA672700
c 29	71.8	0.7	250	14	CA802132
c 30	71.8	0.7	468	12	BM933558
c 31	71.8	0.7	644	29	CNS0220Q
c 32	71.8	0.7	1201	9	ALS13597
c 33	71.6	0.7	1201	13	BX446581
c 34	71.6	0.7	310	10	B3667316
c 35	71.6	0.7	330	14	CK005155
c 36	71.6	0.7	620	9	AV701900
c 37	71.4	0.7	263	14	CD455332
c 38	71.4	0.7	541	14	CA953632
c 39	71.4	0.7	598	9	AV726058
c 40	71.4	0.7	681	12	B517482
c 41	71.4	0.7	978	14	CB196578
c 42	71.4	0.7	1126	13	BX446391
c 43	71.4	0.7	1405	12	BG107670
c 44	71.2	0.7	204	12	BG996016
c 45	71.2	0.7	544	13	BQ186043

ALIGNMENTS

RESULT 1
AV755731/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AV755731 488 bp mRNA linear EST 19-OCT-2000
AV755731 BM Homo sapiens cDNA clone BMFAK803 5', mRNA sequence.
AV755731.1 GI:10913579
EST.
Homo sapiens (human)
Homo sapiens
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 488)
Gu, J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H.,
Gu, Y., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z.,
Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M.,
Lu, G., Yang, Y., Gao, G., Wang, Z., Zhang, Q., Chen, S., Han, Z. and
Chen, Z.
Homo sapiens cDNA BM clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@hgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

Location/Qualifiers
 1. .438
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="BMPAKB03"
 /issue_type="Bone marrow"
 /cell_type="CD34+ hematopoietic stem/progenitor cell"
 /lab_host="BM25.8"
 /clone_lib="BM"
 /note="Vector: pTriplex2; Site_1: sfIIA; Site_2: sfIIB"

ORIGIN

Query Match 1.3%; Score 129; DB 9; Length 488;
 Best Local Similarity 58.9%; Pred. No. 3.4e-10;
 Matches 275; Conservative 0; Mismatches 187; Indels 5; Gaps 3;

QY 1137 GTTGGGACGCTCTTTCTGCTCCGTAATGATAGTGGGGATCTTGGGATCTATTTTC 1196
 Db 472 GTGGTCTACACTCGCTCTGCTGCTCTAGCTCTAGCTTGGGACCTCTGCGACGGAGTGATG 413
 QY 1197 CTGGTCTCCAGCTGTTCACCTTCTGCTCGCGCGCATGACAGACTGCGAGGACTGCAAC 1256
 Db 412 CTTCAGATTGAGTATCA---TCTGGCTTCAGCACCATGAGTTGTGATGATGCAAC 356
 QY 1257 TGCTCAATCTATCCCGGCCATGTATCAGTTCACCGCATG--GCATTGGGATATGATGAA 1315
 Db 355 TGCTCCATCTATCTGGCGGCATCACTGGACACCGTATGAGTACGAGCGACATGATGAA 296
 QY 1316 CTGGTCACTACACAGCCCTAGTGTGTCGAGTTGCTCCGGATCCACACACTCTCGT 1375
 Db 295 CTGGTCTGACACCGCTCTATGATCATGCGGTACGCAATGCGGCTTCTGAGTCTCAT 236
 QY 1376 GGCATATGTTGGCGGGGCGCCACTGGGAGTCTGGGGGCTTGCCTACTATTCCATGGT 1435
 Db 235 AGATATCATACGCGGGCTCACTGGGCGTCAATGTTGCGGTTAGCTTACTTCTATGCA 176
 QY 1436 AGGGAATCGGCTAAAGTTCTGATTTGCGGCTACTCTTTGCCGGCTTACCGGAGAC 1495
 Db 175 GGGAGCGTGGCGAAAGTCTGCTCATCTCTCTGTTGACCGCTGGGCTACACCGAGCAC 116
 QY 1496 CCACAGACGGGAGGTGGCCGCGC-CACACGACTCCGGGTTACGTCCTTTCTCAT 1554
 Db 115 CCAGCTCACAGTGGGCACTGCTAGCGCGCGCCGACAGCTTTCGTCAGCTTCTCACCC 56
 QY 1555 CTGGGCGCTCTAGAAAATCCAGTTGTGAATACCAACGCGAGTGG 1601
 Db 55 CTGGCGCTAAGAGACATCCACTCATTTANGATCTCATATGGG 9

RESULT 2

AV758366/c
 LOCUS AV758366 BM Homo sapiens cDNA clone BMPAKA03 5', mRNA linear EST 19-OCT-2000
 DEFINITION AV758366
 ACCESSION AV758366
 VERSION AV758366.1 GI:10916214
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 492)
 Gu, J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H.,
 Gu, Y., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z.,
 Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M.,
 Lu, G., Yang, Y., Gao, G., Wang, Z., Zhang, Q., Chen, S., Han, Z. and
 Chen, Z.

TITLE Homo sapiens cDNA BM clones
 JOURNAL Unpublished (2000)
 CONTACT: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China

Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES

Location/Qualifiers
 1. .492
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="BMPAKA03"
 /issue_type="Bone marrow"
 /cell_type="CD34+ hematopoietic stem/progenitor cell"
 /lab_host="BM25.8"
 /clone_lib="BM"
 /note="Vector: pTriplex2; Site_1: sfIIA; Site_2: sfIIB"

ORIGIN

Query Match 1.0%; Score 98; DB 9; Length 492;
 Best Local Similarity 55.2%; Pred. No. 2.4e-05;
 Matches 250; Conservative 0; Mismatches 198; Indels 5; Gaps 3;

QY 1153 TCTGCTCCGCTATGACGTGGGGGATCTTGGGATCTATTTTCCTCTCCAGCTGT 1212
 Db 457 TGTGATCAGCTCAGTACGTGTTGGACCTCTGCTTGGGGTATCGCTTGAGCCCACTGA 398
 QY 1213 TCACCTTCTCGCTCGCGGCATGAGACAGTGCAGAGCTGCAACTGCTCAATCTATCCG 1272
 Db 397 TTA---TCTCTCAGCAGCAACATTCGTTTGTGCAAGATGCAACTGCTCATTTCTATCTG 341
 QY 1273 GCATGTATCAGTTC-ACCGATGCTTGGGATATGATGATGAATGCTCAGCTCAACA 1331
 Db 340 GTGATCATCATGAGTACAGTATGCTATGCTATGATGATGATGATGCTGCGCAGCCGT 281
 QY 1332 GCGCTAGTGTCTGTCAGTTGCTCGGATCCCAAGCTGTCGTGGAATGCTGCGGGG 1391
 Db 280 TCATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 221
 QY 1392 GCGCATGCGGAGTCTTGGCGGCTTGGCTACTATTCATGTTAGGAACTGGGCTAAG 1451
 Db 220 GCACACTGCGCGCTCATGTTCGCTCAGCTTACTTCAATGCGAGGAGCGTTGCCCAA 161
 QY 1452 GTTCTGATTTGGCGCTACTCTTTGCGGCGTGTAGCGGGAGACCCACACGCGGGAGG 1511
 Db 160 GTAGCTGTATCATCTCTGTTTCACTCCGCTGGGTTGATGCGAGCAACACGATATCAGG 101
 QY 1512 GTGGGCG-CCACACACCTCGCGGTTACGTCCTCTTCTCATCTGGGCGCTCTCAGAA 1570
 Db 100 ATCACTGCGCGGACGACAGAGTTCGTCAGGTTGCAATGCTGCGCTGCGCTAAGCAGTA 41
 QY 1571 AATCCAGTTTGAATATCAACGCGAGCTGGCA 1603
 Db 40 CATCCAGCTCATGANGCTCATGCGCAATGGNA 8

RESULT 3

BQ232672/c
 LOCUS BQ232672 914 bp mRNA linear EST 02-MAY-2002
 DEFINITION AGENCOURT 7552993 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6058069
 5', mRNA sequence.

ACCESSION BQ232672
 VERSION BQ232672.1 GI:20414072
 SOURCE EST.
 ORGANISM Homo sapiens (human)

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 914)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgagbs-r@mail.nih.gov
 Tissue Procurement: DCTB/DTF/Gazdar

```
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: ILM10020 row: i column: 14
High quality sequence stop: 590.
Location/Qualifiers
source
1. .914
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6058069"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_68"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
ORIGIN
Query Match 0.8%; Score 77.8; DB 13; Length 914;
Best Local Similarity 49.6%; Pred. No. 0.031;
Matches 112; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
QY 9297 GCGGACCCCGTGGTTCGGTGGCTACTCTCTACTCTCTAGGGGTAGGCAATTAC 9356
Db 694 GCGCCCCCGGGGGATTTTTTTTTCGCGCCCCCTTTNNNNNNNNNNNNNNNN 635
QY 9357 CTGCTCCCCAACGATGAACGGGGAGCTAACCACTCCAGGCGCTTAAGCAATTCCTGTT 9416
Db 634 TTTTNTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 575
QY 9417 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 9476
Db 574 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 515
QY 9477 TTTCTTTTCCCTCTTTTAAAGGGGCTCATCTTAGCCCTAGCA 9522
Db 514 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 469
RESULT 4
BG108244 807 bp mRNA linear EST 30-JAN-2001
LOCUS 602280129F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4367719 5',
DEFINITION mRNA sequence.
ACCESSION BG108244
VERSION BG108244.1 GI:12602010
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 807)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: ILM10020 row: j column: 08
High quality sequence stop: 565.
Location/Qualifiers
source
1. .807
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4367719"
/tissue_type="osteosarcoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_86"
/notes="Organ: bone; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.533 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
ORIGIN
Query Match 0.8%; Score 76.6; DB 12; Length 807;
Best Local Similarity 64.2%; Pred. No. 0.049;
Matches 115; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 9316 CGTGTGCTACTCCACTTCTGTAGGGGTAGGCAATTTACCTGTCGCCAACCGATGAA 9375
Db 638 CGTGTGCCCCGCTTTACTACGATGCTGTTGGCTATTGCTCCCTCGTGTGTTTA 579
QY 9376 CGGGAGCTAACCACTCCAGGCGCTTAAGCAATTCCTGTTTNTTNTTNTTNTT 9435
Db 578 ATTGCGGGCGGCCCCCTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 519
QY 9436 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 9494
Db 518 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 460
RESULT 5
CNS011GK 796 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence 17 end of BAC
DEFINITION BACN06M18 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL100238
VERSION AL100238.1 GI:5611849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Epiphytodea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 796)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaut at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBel0BAC11.
FEATURES
source
1. .796
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN06M18"
/plasmid="pBel0BAC11"
/notes="end : T7"
ORIGIN
Query Match 0.8%; Score 76; DB 29; Length 796;
Best Local Similarity 50.5%; Pred. No. 0.061;
Matches 103; Conservative 28; Mismatches 73; Indels 0; Gaps 0;
```

```

1. .548
location,qualifiers
/organism="Danio rerio"
/mol_type="mRNA"
/strain="SJD"
/db_xref="taxon:7955"
/clone="IWAGE:5619999"
/tissue_type="embryo, day 3"
/lab_host="DH103"
/clone_lib="zebrafish SJD day 3 embryo"
/notes="vector: PAMPI; Site 1: EcoRI; Site 2: NotI; First
strand cDNA synthesis was primed using oligo-dT on
magnetic beads with an additional primer
5'-ggcgcgctaatcagctacta-tagg-3'. Second strand
synthesis was a 3-cycle PCR using the primers
5'-ggcgcgtaaacactcactag-3' and
5'-aagcagggtaacagcagagacttt-ttttttttttn-3'. cDNA
was subsequently amplified in a 7-cycle PCR with the
following primers: 5'-ggcgcgtaacactcactag-3', and
5'-aagcaggggt-aacaacgacg. Deoxy-UMP adaptors were added in

```

[illegible]

<http://image.ihm.gov>
This clone was contributed by the Brain Molecular Anatomy Project (EMAP)
The following repetitive elements were found in this cDNA
sequence: 428-543, >POLY_A#simple repeat
seg primer: PYX-5.

```

1. .562
location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5712969"
/tissue_type="whole brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NH BMAP PA0"
/note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 5:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an

```


source

```

1. .202
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CI0186"
/note="Organ: colon; ins; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

ORIGIN

```

Query Match      0.8%; Score 73; DB 10; Length 202;
Best Local Similarity 65.8%; Pred. No. 0.23; 55; Indels 0; Gaps 0;
Matches 106; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 9352 TTTACCTGCTCCCAACCGATGAACGGGAGCTAACCACTCCAGGCTTAAGCCATTTC 9411
      |||||
Db 187 TTTCCCAACGCCGCCCTTTAATATTCGCCGCCGCCGCCGCCGCCGCCGCCCTTTT 128
      |||||

QY 9412 TGTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9471
      |||||
Db 127 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 68
      |||||

QY 9472 TTTCCCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9512
      |||||
Db 67 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 27
      |||||

```

RESULT 15

```

AL643867/c
LOCUS      558 bp mRNA linear EST 07-NOV-2001
DEFINITION AL643867 XGC-gastrula Silurana tropicalis cDNA clone L1G4f7 5',
            mRNA sequence.
ACCESSION  AL643867
VERSION     AL643867.1 GI:16795992
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
            Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1 (bases 1 to 558)
            Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
            Sanger Xenopus tropicalis EST project 2001 (10_2001)
            Unpublished (2001)
            Contact: Huckle E
            Sanger Institute
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk

```

```

Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: L1G4f7.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers

```

FEATURES

```

1. .558
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="L1G4f7"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dT primed from 5ug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

```

ORIGIN

```

Query Match      0.8%; Score 73; DB 9; Length 558;
Best Local Similarity 65.8%; Pred. No. 0.19;
Matches 106; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 9334 TTTCTGTAGGGGTAGGCATTTACCTGCTCCCAACCGATGAACGGGAGCTAACCACTCC 9393
      |||||
Db 411 TTTAAGTTTGGGTTTTCGTAAAAACACCCGCCCATGAAGGGGAAGGGGTTAAAGCCCTTC 352
      |||||

QY 9394 AGGCCTTAAGCCATTTCTCTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 9453
      |||||
Db 351 GGCCTCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 292
      |||||

QY 9454 CTTTCCTTTCTCTTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9494
      |||||
Db 291 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 251
      |||||

```

Search completed: April 8, 2004, 21:52:34
Job time : 20680 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: April 9, 2004, 05:57:21 ; Search time 3212 Seconds
(without alignments)
12690.368 Million cell updates/sec

Title: US-09-662-454-4
Perfect score: 9595
Sequence: 1 gccagccctctgagggsc.....ggcctctgagatcatgt 9595

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001as.*
- 5: Geneseq2001bs.*
- 6: Geneseq2002s.*
- 7: Geneseq2003as.*
- 8: Geneseq2003bs.*
- 9: Geneseq2003cs.*
- 10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9595	100.0	9595	2	AAQ24843
2	9595	100.0	9595	4	AAQ24843
3	9593.4	100.0	9595	4	AAQ24843
4	8311.8	86.6	9413	6	AAQ24843
5	8311.8	86.6	9413	7	AAQ24843
6	8310.2	86.6	9413	7	AAQ24843
7	8308.6	86.6	9413	2	AAQ24843
8	8307	86.6	9413	2	AAQ24843
9	8294.2	86.4	9413	2	AAQ24843
10	8250.2	86.0	9605	6	AAQ24843
11	8248.6	86.0	9605	6	AAQ24843
12	8248.6	86.0	9605	6	AAQ24843
13	8248.6	86.0	9605	6	AAQ24843
14	8248.6	86.0	9605	6	AAQ24843
15	8248.6	86.0	9605	6	AAQ24843
16	8248.6	86.0	9605	6	AAQ24843
17	8248.6	86.0	9605	6	AAQ24843
18	8248.6	86.0	9605	6	AAQ24843
19	8248.6	86.0	9605	6	AAQ24843
20	8247	86.0	9605	6	AAQ24843
21	8237.2	85.8	9608	6	AAQ24843
22	8219.6	85.7	9436	2	AAQ24843
23	8170.4	85.2	9416	2	AAQ24843

24	8170.4	85.2	9416	2	AAQ21829
25	8168.8	85.1	9416	2	AAQ21829
26	8099.8	84.4	9472	2	AAQ33282
27	8067	84.1	9609	6	AAQ33038
28	8051.4	83.9	9405	2	AAQ40426
29	8032.8	83.7	9391	2	AAQ64175
30	8021.6	83.6	9391	2	AAQ38959
31	7946	82.8	9402	2	AAQ41345
32	7918.6	82.5	11076	3	AAQ98965
33	7914	82.5	11062	6	AAQ25331
34	7201.4	75.1	9599	2	AAQ24833
35	6612	68.9	7911	2	AAQ32436
36	6335.8	66.0	9599	2	AAQ24832
37	6335.8	66.0	9599	4	AAQ66938
38	6334.2	66.0	9599	4	AAQ23491
39	6264	65.3	12980	2	AAQ59364
40	6264	65.3	12980	7	AAQ62469
41	6262.4	65.3	12980	6	ABX87286
42	6258	65.2	9622	7	AAQ54424
43	6254.8	65.2	9646	2	AAQ59361
44	6254.8	65.2	9646	6	ABX87285
45	6254.8	65.2	9646	7	AAQ62466

ALIGNMENTS

RESULT 1

AAQ24843

ID AAX24843 standard; DNA; 9595 BP.

AC AAX24843;

XX

XX

DT 21-JUN-1999 (first entry)

XX

DE Infectious hepatitis C virus genotype 1b strain HC-J4 genome.

XX HCV; infectious clone; infection; diagnosis; therapy; vaccine; screening;

KW HCV; antiviral; virucide; ss.

XX

OS Hepatitis C virus.

XX

FH Key Location/Qualifiers

FT CDS 342..9374

FT /*tag= a

XX

XX WO9904008-A2.

XX

XX 28-JAN-1999.

XX

XX 16-JUL-1998; 98WO-US014688.

XX

XX 18-JUL-1997; 97US-0053062P.

XX

XX 27-JAN-1998; 98US-00014416.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA Yanagi M, Bukh J, Emerson SU, Purcell RH;

PI WPI; 1999-132252/11.

XX P-PSDB; AAW98022.

XX

XX New isolated hepatitis C virus nucleic acids - used to develop products

PT for the diagnosis, prevention and treatment of HCV infections and for

PT developing screening assays.

XX Claim 3; Fig 14A-F; 126pp; English.

XX

XX The present sequence comprises the nucleic acid sequence of the genome of

CC infectious hepatitis C virus (HCV) genotype 1b strain HC-J4 (ATCC 209596)

CC that is capable of expressing this virus when transfected into cells. HC-

CC J4 was obtained from acute phase plasma of a chimpanzee experimentally

CC infected with serum containing HC-J4/91. The claimed infectious nucleic

1861 ATGTTTACCCCAAGCCCTGTTGTTGGTGGGACCAACGATCGTTCGGGTGTCCTTACGT 1920
 1861 ATGTTTACCCCAAGCCCTGTTGTTGGTGGGACCAACGATCGTTCGGGTGTCCTTACGT 1920
 1921 ATAGCTGGGGGAGATGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
 1921 ATAGCTGGGGGAGATGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
 1981 GCAACTGGTTCGGCTGTACATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
 1981 GCAACTGGTTCGGCTGTACATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
 2041 CCCCGTGTAACATCGGGGGGTCGGTAAACGACCTTGTATCGCCACAGAGAGAGAGAG 2100
 2041 CCCCGTGTAACATCGGGGGGTCGGTAAACGACCTTGTATCGCCACAGAGAGAGAGAG 2100
 2101 GGAAGACCCCGAGGCTACTTACAAAATGTGGTGGGGGCTGTTGATGATGATGATGATG 2160
 2101 GGAAGACCCCGAGGCTACTTACAAAATGTGGTGGGGGCTGTTGATGATGATGATGATG 2160
 2161 GCTTGTAGTACTACCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTTCATCT 2220
 2161 GCTTGTAGTACTACCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTTCATCT 2220
 2221 TTAAGGTTAGGATGTATGTGGGGGCTGGAGACAGGCTCAATGCCGATGCAATTGGA 2280
 2221 TTAAGGTTAGGATGTATGTGGGGGCTGGAGACAGGCTCAATGCCGATGCAATTGGA 2280
 2281 CTCGAGGAGCGCTGTAATCTGGAGGACAGGATAGGTGAGAGAGAGAGAGAGAGAGAG 2340
 2281 CTCGAGGAGCGCTGTAATCTGGAGGACAGGATAGGTGAGAGAGAGAGAGAGAGAGAG 2340
 2341 TGCTACACAGAGTGCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400
 2341 TGCTACACAGAGTGCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400
 2401 CTGGTTTGTATCTCCATCAGAAATCGTGAGCGTGAATACCTGACGCTGTAGGGT 2460
 2401 CTGGTTTGTATCTCCATCAGAAATCGTGAGCGTGAATACCTGACGCTGTAGGGT 2460
 2461 CAGCGTTGTCTCCTTTGCAATCAATGGGAGTACAACCTGTTGCTTTTCTTCTCTG 2520
 2461 CAGCGTTGTCTCCTTTGCAATCAATGGGAGTACAACCTGTTGCTTTTCTTCTCTG 2520
 2521 CAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
 2521 CAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
 2581 CCTTAGAGAACTTGGTGTCTCAATCGGCGTCCGTCGGCGGAGCGATGGTATCTCT 2640
 2581 CCTTAGAGAACTTGGTGTCTCAATCGGCGTCCGTCGGCGGAGCGATGGTATCTCT 2640
 2641 CTTTCTGTGTTCTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
 2641 CTTTCTGTGTTCTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
 2701 CGTATGCTTTTATGGGATGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760
 2701 CGTATGCTTTTATGGGATGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760
 2761 CTTACGCTTTGGACCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820
 2761 CTTACGCTTTGGACCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820
 2821 TATTTCTGACTTGTACACATACAAAAGTGTTCCTACCTAGGCTCATATGTTGTTTAC 2880
 2821 TATTTCTGACTTGTACACATACAAAAGTGTTCCTACCTAGGCTCATATGTTGTTTAC 2880
 2881 AATACTTTATCAGAGCGCGGAGGACATGAGAGTGTGGTTCCTGCTGCTGCTGCTGCTGCT 2940
 2881 AATACTTTATCAGAGCGCGGAGGACATGAGAGTGTGGTTCCTGCTGCTGCTGCTGCTGCT 2940

2941 GGGAGGCGCGCATGCCATCATCTCCTCACGTGCGGTTTCATCCAGAGTTAAATTTTG 3000
 2941 GGGAGGCGCGCATGCCATCATCTCCTCACGTGCGGTTTCATCCAGAGTTAAATTTTG 3000
 3001 ACATACOMAACTCTGCTGCGCATCTCGGCCCTCATGCTGCTCAGGCTGCGCAAA 3060
 3001 ACATACOMAACTCTGCTGCGCATCTCGGCCCTCATGCTGCTCAGGCTGCGCAAA 3060
 3061 CGAGAGTGGCTACTTCTGTCGCGCTCAAGGGCTCATCTGTCATGCTGCTTGTAGTGGAA 3120
 3061 CGAGAGTGGCTACTTCTGTCGCGCTCAAGGGCTCATCTGTCATGCTGCTTGTAGTGGAA 3120
 3121 AAGTGGCGGGGCTCATCTATGTCCTTCAAGTGGTCTTCAAGTGGCTGCTGAGAGTA 3180
 3121 AAGTGGCGGGGCTCATCTATGTCCTTCAAGTGGTCTTCAAGTGGCTGCTGAGAGTA 3180
 3181 CGTAGCTTTATTAACCATCTTACCCCACTGCGGGAATGGGCCCAACGCGGCTTACGAGAC 3240
 3181 CGTAGCTTTATTAACCATCTTACCCCACTGCGGGAATGGGCCCAACGCGGCTTACGAGAC 3240
 3241 TTGCGGTGGCGTAGAGCCGCTGCTTCTCGCCCATGGAGACCAAGGTTCATCACTGGG 3300
 3241 TTGCGGTGGCGTAGAGCCGCTGCTTCTCGCCCATGGAGACCAAGGTTCATCACTGGG 3300
 3301 GAGCAGACAACCGCTGCTGCTGGGACATCATCTTGGTCTAACCCTTCCGCCGAGGG 3360
 3301 GAGCAGACAACCGCTGCTGCTGGGACATCATCTTGGTCTAACCCTTCCGCCGAGGG 3360
 3361 GGAAGAGATATTTTGGGACCGGCTGATGCTCGAAGGCAAGGTGGCCACTCTTGG 3420
 3361 GGAAGAGATATTTTGGGACCGGCTGATGCTCGAAGGCAAGGTGGCCACTCTTGG 3420
 3421 CGCCCATCAACGCTTACTCCCAACAAACGCGGGGCTACTTGGTTCATCATCACTAGCC 3480
 3421 CGCCCATCAACGCTTACTCCCAACAAACGCGGGGCTACTTGGTTCATCATCACTAGCC 3480
 3481 TCACAGCGCGGACAAAGAACAGGTCGAAGGGAGGTTCAAGTGGTTCATCCGAAACAC 3540
 3481 TCACAGCGCGGACAAAGAACAGGTCGAAGGGAGGTTCAAGTGGTTCATCCGAAACAC 3540
 3541 AATCTTCTGCGGACCTGATCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
 3541 AATCTTCTGCGGACCTGATCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
 3601 CGAAGACCTTAGCGGTCGCAAGGTCCTCAATCACCAAAATGTACCAATGTAGACCTGG 3660
 3601 CGAAGACCTTAGCGGTCGCAAGGTCCTCAATCACCAAAATGTACCAATGTAGACCTGG 3660
 3661 AACTCTGCTGGCAGGCG 3720
 3661 AACTCTGCTGGCAGGCG 3720
 3721 GCTCGGACCTTTACTTTGTCAGAGACATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 3780
 3721 GCTCGGACCTTTACTTTGTCAGAGACATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 3780
 3781 ACAGCAGGCGGAGTCTACTCTTCCCGCAGGCGCTTCTCTACCTGAAAGGCTCCTCGGCTG 3840
 3781 ACAGCAGGCGGAGTCTACTCTTCCCGCAGGCGCTTCTCTACCTGAAAGGCTCCTCGGCTG 3840
 3841 GTCCATTTGCTTTCCTTCCGGGACGCTGCTGGGCTCTTCCGGGCTGCTGCTGCTGCTGCT 3900
 3841 GTCCATTTGCTTTCCTTCCGGGACGCTGCTGGGCTCTTCCGGGCTGCTGCTGCTGCTGCT 3900
 3901 GGGGGGTCCGGAAGGCGGTGAGCTTCAATACCGCTTGTAGTCTATGGAATACCATCGGT 3960
 3901 GGGGGGTCCGGAAGGCGGTGAGCTTCAATACCGCTTGTAGTCTATGGAATACCATCGGT 3960
 3961 CTCGGCTTCTCAGACAACCTCAACCCCGGCTGTACCGGACACATTCACAGTGGCAC 4020
 3961 CTCGGCTTCTCAGACAACCTCAACCCCGGCTGTACCGGACACATTCACAGTGGCAC 4020
 4021 ATCTGACGCTCTCTACTGCGAGCGGACAGAGACCAAGTGGCGGCTGCTGCTGCTGCTGCT 4080

[illegible]

5101	TGGCATATCAAGCTACAGTGTGGCCACAGGCTCAAGCTCCACTCCATCTGTTGGGACACAA	5161
5161	TGTGGAAGTGTCTCATACGGCTGAACACTACATCTGACGGGCAACACCCCTGCTGTATA	5221
5161	TGTGGAAGTGTCTCATACGGCTGAACACTACATCTGACGGGCAACACCCCTGCTGTATA	5220
5221	GGCTAGGAGCGCTCCAAAATGAGGTATCTCTCACACACCCCATACTAAATACATCATGG	5280
5221	GGCTAGGAGCGCTCCAAAATGAGGTATCTCTCACACACCCCATACTAAATACATCATGG	5280
5281	CATGATCTCGGCTGACCTGGAGGTCGTCACTAGCAGCTGGTGTCTGCTAGGCGGAGTCC	5340
5281	CATGATCTCGGCTGACCTGGAGGTCGTCACTAGCAGCTGGTGTCTGCTAGGCGGAGTCC	5340
5341	TTGACAGCTTTGGCTCGCATATCTGCTTGACGACAGGCAAGTGTGTATTTGTGGGCAAGATCA	5400
5341	TTGACAGCTTTGGCTCGCATATCTGCTTGACGACAGGCAAGTGTGTATTTGTGGGCAAGATCA	5400
5401	TCTTGTCCGGGAAGCAGCTGTGCTTCCGACAGGGAAGTCTCTTACACAGGAGTTCGATG	5460
5401	TCTTGTCCGGGAAGCAGCTGTGCTTCCGACAGGGAAGTCTCTTACACAGGAGTTCGATG	5460
5461	AGATGGAAGTGTGCTTCACAACTTCTTACATCGAGCAGGGAATGCAGTTCGCCGAGC	5520
5461	AGATGGAAGTGTGCTTCACAACTTCTTACATCGAGCAGGGAATGCAGTTCGCCGAGC	5520
5521	AATTCAGCAAAAGCGCTGGGTGTGTGCAGACGCCACCAAGCAGCAGCGGAGGCTGTG	5580
5521	AATTCAGCAAAAGCGCTGGGTGTGTGCAGACGCCACCAAGCAGCAGCGGAGGCTGTG	5580
5581	CTCCGCTGTGAGTCCAAAGTGGCAGCCCTTGAGACCTTCTGGGCGAAGCACATGTGGA	5640
5581	CTCCGCTGTGAGTCCAAAGTGGCAGCCCTTGAGACCTTCTGGGCGAAGCACATGTGGA	5640
5641	ATTTCATCAGCGGAATACAGTAACCTAGCAGGCTTATCCACTCTGCTCGGAAACCCCGCA	5700
5641	ATTTCATCAGCGGAATACAGTAACCTAGCAGGCTTATCCACTCTGCTCGGAAACCCCGCA	5700
5701	TAGCATCATGATGGCATTTACAGCTTCTATCACTAGCCGCTCAACCCCAACCAACCCC	5760
5701	TAGCATCATGATGGCATTTACAGCTTCTATCACTAGCCGCTCAACCCCAACCAACCCC	5760
5761	TCTGTTTAAACATCTTGGGGGATPGGTGTCTGCCAAATGCTCTCTCCACGCGTGGT	5820
5761	TCTGTTTAAACATCTTGGGGGATPGGTGTCTGCCAAATGCTCTCTCCACGCGTGGT	5820
5821	CAGCTTTGTGGGCGCGGATCCCGAGCGGCTCTTGGCAGCATAGGCTTGGGAAGG	5880
5821	CAGCTTTGTGGGCGCGGATCCCGAGCGGCTCTTGGCAGCATAGGCTTGGGAAGG	5880
5881	TGCTGTGGAACATCTTGGCGGGCTATGGGCGAGGTTAGCCGCGCATCTGTGCGCTTTA	5940
5881	TGCTGTGGAACATCTTGGCGGGCTATGGGCGAGGTTAGCCGCGCATCTGTGCGCTTTA	5940
5941	AGGTCAATGAGCGGCGAGTGCCTTCACCGAGCACTGTGTCACTTCTCCCTGCCATCC	6000
5941	AGGTCAATGAGCGGCGAGTGCCTTCACCGAGCACTGTGTCACTTCTCCCTGCCATCC	6000
6001	TCTCTCTGTGCTCCTGTGCTCGGGTCTGTGTCGACGCAATACTGCTCGGCACTGG	6060
6001	TCTCTCTGTGCTCCTGTGCTCGGGTCTGTGTCGACGCAATACTGCTCGGCACTGG	6060
6061	GGCCGGGAGAGGGGGCTGTGCAGTGAATGAACCGGCTGATACGTTTCGTTTCGGGGGTA	6120
6061	GGCCGGGAGAGGGGGCTGTGCAGTGAATGAACCGGCTGATACGTTTCGTTTCGGGGGTA	6120
6121	ACCAAGTCTCCCTACGCACTATGTGCTTGAGAGCAAGCTTCAGCACTGTCTCAGTA	6180
6121	ACCAAGTCTCCCTACGCACTATGTGCTTGAGAGCAAGCTTCAGCACTGTCTCAGTA	6180
6181	TCTCTCTAGGCTTACCATCACTCAACTGCTGAAGCGGCTCCACAGTGGATTATAGG	6240
6181	TCTCTCTAGGCTTACCATCACTCAACTGCTGAAGCGGCTCCACAGTGGATTATAGG	6240

QY 6241 ACTGCTCTACGCATCTCCGCTCTGCTGCTAGGATGTTTGGGATCGATATGCAAG 6300
 Db 6241 ACTGCTCTACGCATCTCCGCTCTGCTGCTAGGATGTTTGGGATCGATATGCAAG 6300
 QY 6301 TGTGACTGACTCAAGACCTGGCTCCAGTTCAGATCCCTCCGCGGTTACCGGGAGTCC 6360
 Db 6301 TGTGACTGACTCAAGACCTGGCTCCAGTTCAGATCCCTCCGCGGTTACCGGGAGTCC 6360
 QY 6361 CTCTTCCTGTCATCCCAACCGGGTACAGGGAGTCTGGCGGGGACGCTATCGATAA 6420
 Db 6361 CTCTTCCTGTCATCCCAACCGGGTACAGGGAGTCTGGCGGGGACGCTATCGATAA 6420
 QY 6421 CCACCTGSCCATCGGAGACAGATCGCGGACATGTCAAAACGGTTCATGAGATCG 6480
 Db 6421 CCACCTGSCCATCGGAGACAGATCGCGGACATGTCAAAACGGTTCATGAGATCG 6480
 QY 6481 TAGGGCTGAACTCGAGCAACAGTGGGACGGAAGCTTCCCATCAACGATACCA 6540
 Db 6481 TAGGGCTGAACTCGAGCAACAGTGGGACGGAAGCTTCCCATCAACGATACCA 6540
 QY 6541 CGGGACCTTGACACACCTCCCGGCGCCCACTATTCCAGGGGCTATGGCGGGTGGCTG 6600
 Db 6541 CGGGACCTTGACACACCTCCCGGCGCCCACTATTCCAGGGGCTATGGCGGGTGGCTG 6600
 QY 6601 CTGAGGAGTACGTGGAGGTTACCGGTGTGGGGGATTTCCACTACGTGACGGGATGACCA 6660
 Db 6601 CTGAGGAGTACGTGGAGGTTACCGGTGTGGGGGATTTCCACTACGTGACGGGATGACCA 6660
 QY 6661 CTGACACGTAAAGTGCCTATGCGAGTTCGGGCGCCCGAATCTTTCACGGAGTGGATG 6720
 Db 6661 CTGACACGTAAAGTGCCTATGCGAGTTCGGGCGCCCGAATCTTTCACGGAGTGGATG 6720
 QY 6721 GAGTGGCTGTGACAGTACCGCTCCGGCGTGAACCTCTTTCACGGAGGAGCGTCACT 6780
 Db 6721 GAGTGGCTGTGACAGTACCGCTCCGGCGTGAACCTCTTTCACGGAGGAGCGTCACT 6780
 QY 6781 TCCAGTTCGGGCTCAACCAATACTTGGTTCGGGTTCGAGCTCCCATGCGAGCCCGAACCGG 6840
 Db 6781 TCCAGTTCGGGCTCAACCAATACTTGGTTCGGGTTCGAGCTCCCATGCGAGCCCGAACCGG 6840
 QY 6841 ACCTACAGTGTCTTCTCCATGCTCACCGATCCCTCCCGACATACAGCAGACGCGTA 6900
 Db 6841 ACCTACAGTGTCTTCTCCATGCTCACCGATCCCTCCCGACATACAGCAGACGCGTA 6900
 QY 6901 AGCGTAGGCTGTAGAGGTCTCCCGCTCTTTAGCCAGCTCATCAGCTAGCCAGTTGT 6960
 Db 6901 AGCGTAGGCTGTAGAGGTCTCCCGCTCTTTAGCCAGCTCATCAGCTAGCCAGTTGT 6960
 QY 6961 CTGCGCTCTTTTGAAGGCGACATGCACTACCCACCATGACTCCCGGACGCTGACCTCA 7020
 Db 6961 CTGCGCTCTTTTGAAGGCGACATGCACTACCCACCATGACTCCCGGACGCTGACCTCA 7020
 QY 7021 TCGAGGCAACCTCTTGTGCGGAGAGATGGGCGGAAACATCATCTCGGCTGGAGTTCAG 7080
 Db 7021 TCGAGGCAACCTCTTGTGCGGAGAGATGGGCGGAAACATCATCTCGGCTGGAGTTCAG 7080
 QY 7081 AGAATAGGTAGTAATCTCTGACTCTTTTGAACCGGTTCACGGGGGGGATGAGAGG 7140
 Db 7081 AGAATAGGTAGTAATCTCTGACTCTTTTGAACCGGTTCACGGGGGGGATGAGAGG 7140
 QY 7141 AGATATCCGTCGCGGGAGATCTGCGAAATCCAGGAAGTTTCCCTCAGCGTTGCCCA 7200
 Db 7141 AGATATCCGTCGCGGGAGATCTGCGAAATCCAGGAAGTTTCCCTCAGCGTTGCCCA 7200
 QY 7201 TATGGGACGCGCGGACTACAACTCTCTGACTCTTTTGAACCGGTTCACGGGGGGGATGAGAGG 7260
 Db 7201 TATGGGACGCGCGGACTACAACTCTCTGACTCTTTTGAACCGGTTCACGGGGGGGATGAGAGG 7260
 QY 7261 TCCTTCGGTGTGACAGGATGCCATTGCACTACCAAGGCTCTCCCAATACCACTC 7320
 Db 7261 TCCTTCGGTGTGACAGGATGCCATTGCACTACCAAGGCTCTCCCAATACCACTC 7320

QY 7321 CACGAGAAAGAGCAGCGTGTGCTGACAGAAATCCAAATGTCTGTCTGCTTGGCGGAGC 7380
 Db 7321 CACGAGAAAGAGCAGCGTGTGCTGACAGAAATCCAAATGTCTGTCTGCTTGGCGGAGC 7380
 QY 7381 TCGCCACTAAGACCTTCCGCTAGCTCCGATCGTCCGCTTGTATAGCGGACGCGGACCG 7440
 Db 7381 TCGCCACTAAGACCTTCCGCTAGCTCCGATCGTCCGCTTGTATAGCGGACGCGGACCG 7440
 QY 7441 CCCTTCTGACCTGCGCTCCGACGAGGTGACAAAGATCCGAGTTCGATGCTGCTCTCT 7500
 Db 7441 CCCTTCTGACCTGCGCTCCGACGAGGTGACAAAGATCCGAGTTCGATGCTGCTCTCT 7500
 QY 7501 CCATGCCCGCTTGAAGGGGAGCGCGGGACCCCGATCTCAGCGACCGGTTTGGTCTA 7560
 Db 7501 CCATGCCCGCTTGAAGGGGAGCGCGGGACCCCGATCTCAGCGACCGGTTTGGTCTA 7560
 QY 7561 CCGTAGTAGAGAGCTAGTAGGATGCTGCTGCTGCTCAATGTCTTATAGTGGACAG 7620
 Db 7561 CCGTAGTAGAGAGCTAGTAGGATGCTGCTGCTGCTCAATGTCTTATAGTGGACAG 7620
 QY 7621 GCGCCTGATCAGCATGCGCTCGGAGGAAAGTAAAGTGCCTCAACCCGTTGAGCA 7680
 Db 7621 GCGCCTGATCAGCATGCGCTCGGAGGAAAGTAAAGTGCCTCAACCCGTTGAGCA 7680
 QY 7681 ACTCTTGTGCTGCTCAACAACTAGTGTCTAGCCCAACATCCCGGACGCAAGCTCC 7740
 Db 7681 ACTCTTGTGCTGCTCAACAACTAGTGTCTAGCCCAACATCCCGGACGCAAGCTCC 7740
 QY 7741 GCGAGAGAGTCACTTGTGACAGATTCGAGTCTGATGATCAATACGCGGACGTTAC 7800
 Db 7741 GCGAGAGAGTCACTTGTGACAGATTCGAGTCTGATGATCAATACGCGGACGTTAC 7800
 QY 7801 TCAAGAGATGAAGCGAAGCGGTCCACAGTTAAGCTTAAGCTTCTATATAGAGGAGG 7860
 Db 7801 TCAAGAGATGAAGCGAAGCGGTCCACAGTTAAGCTTAAGCTTCTATATAGAGGAGG 7860
 QY 7861 CTGCAAGCTGACGCGCCCACTTCGGCCAAATCCAAATTTGGCTATGGGCAAGGACG 7920
 Db 7861 CTGCAAGCTGACGCGCCCACTTCGGCCAAATCCAAATTTGGCTATGGGCAAGGACG 7920
 QY 7921 TCGGAACTTCCAGCAGGCGGCTTAACCAATCCGCTCCGCTGGGAGGACTTCTGCTG 7980
 Db 7921 TCGGAACTTCCAGCAGGCGGCTTAACCAATCCGCTCCGCTGGGAGGACTTCTGCTG 7980
 QY 7981 AAGACACTGAAACACCAATTTGACACCAATCATGCGMAAAAGTGAAGTTTCTGCTCC 8040
 Db 7981 AAGACACTGAAACACCAATTTGACACCAATCATGCGMAAAAGTGAAGTTTCTGCTCC 8040
 QY 8041 AACAGAGAGGAGCGCGCAAGCAGCTCGCTTATCGTATTTCCAGACCTGGGAGTTTC 8100
 Db 8041 AACAGAGAGGAGCGCGCAAGCAGCTCGCTTATCGTATTTCCAGACCTGGGAGTTTC 8100
 QY 8101 GTGTATGCGAGAGATGCGGCTTTAGCAGTGTCTCCACCTTCTCAGGCGCTGATCG 8160
 Db 8101 GTGTATGCGAGAGATGCGGCTTTAGCAGTGTCTCCACCTTCTCAGGCGCTGATCG 8160
 QY 8161 GCTCTCTCATACGGAATTTCAATACTCCCGCAAGCAGCGGTGAGTTCTCTGTTGAATACCT 8220
 Db 8161 GCTCTCTCATACGGAATTTCAATACTCCCGCAAGCAGCGGTGAGTTCTCTGTTGAATACCT 8220
 QY 8221 GGAATCAAAGAAATGCCCTATGCGGCTTCTCATATGACACCGCTGTTTTGACTCAACGG 8280
 Db 8221 GGAATCAAAGAAATGCCCTATGCGGCTTCTCATATGACACCGCTGTTTTGACTCAACGG 8280
 QY 8281 TCACTAGAGTACATTCGTTGAGGAGTCAATTTACCAATGTTGTGACTTGGCCCCCG 8340
 Db 8281 TCACTAGAGTACATTCGTTGAGGAGTCAATTTACCAATGTTGTGACTTGGCCCCCG 8340
 QY 8341 AGGCCAGACGCGCTAAGGTGCTCAGAGCGGCTTTACATCGGGGCTCCCTGACTA 8400
 Db 8341 AGGCCAGACGCGCTAAGGTGCTCAGAGCGGCTTTACATCGGGGCTCCCTGACTA 8400
 QY 8401 ACTCAAAGGCGAGAACTGCGGTTATTCGCGGTTGCGCGCAAGTGGCGTGTGACTA 8460

8401	Db	ACTCAAAAGGCGAGAACTGGCGTTATCGCGGTTCGCGCGCAGTGGCGGTCTGACACATTA	8460
8461	Qy	GCTGGGTAATACCCCTCAGATGTTACTTGAAGGCCACTGCAGCTCTGTCGAGTGTGCAAAAC	8520
8461	Db	GCTGGGTAATACCCCTCAGATGTTACTTGAAGGCCACTGCAGCTCTGTCGAGTGTGCAAAAC	8520
8521	Qy	TCCAGGACTGCACGATGCTCGTGAACGGAGACGACCTTGTGTTATCTGTGAAGCGCGG	8580
8521	Db	TCCAGGACTGCACGATGCTCGTGAACGGAGACGACCTTGTGTTATCTGTGAAGCGCGG	8580
8581	Qy	GAACCCAGGAGGATCGGCGGCCCTACGAGCTTCAGGAGGCTATGACTAGTATTCG	8640
8581	Db	GAACCCAGGAGGATCGGCGGCCCTACGAGCTTCAGGAGGCTATGACTAGTATTCG	8640
8641	Qy	CCCCCCCCGGGGATCGCCCCAACCAAGATACGACCTGGAGCTGATAACATCATGTTCCT	8700
8641	Db	CCCCCCCCGGGGATCGCCCCAACCAAGATACGACCTGGAGCTGATAACATCATGTTCCT	8700
8701	Qy	CGAATGTTCAGTTCGGCAGATGCATCTGGCAAAAGGTATACTACTCACCCTGAC	8760
8701	Db	CGAATGTTCAGTTCGGCAGATGCATCTGGCAAAAGGTATACTACTCACCCTGAC	8760
8761	Qy	CCACCAACCCCTTGCAACGGGCTCGGTGGAGACAGCTAGACACATCCCAATCAATCTT	8820
8761	Db	CCACCAACCCCTTGCAACGGGCTCGGTGGAGACAGCTAGACACATCCCAATCAATCTT	8820
8821	Qy	GGCTAGGCAATATCATATGATGCGGCCACCTATGGCAAGGATGATCTCATGACTC	8880
8821	Db	GGCTAGGCAATATCATATGATGCGGCCACCTATGGCAAGGATGATCTCATGACTC	8880
8881	Qy	ACTTTTCTCCATCCTTCTAGCTCAAGAGCAACTTGAAAAAGCCCTGGATGTGCATCT	8940
8881	Db	ACTTTTCTCCATCCTTCTAGCTCAAGAGCAACTTGAAAAAGCCCTGGATGTGCATCT	8940
8941	Qy	ACGGGGTTCGTACTCCATTTGAGCCACTGCACTTCTCAGATCATTTGAACGACTCCATG	9000
8941	Db	ACGGGGTTCGTACTCCATTTGAGCCACTTGTGACCTTCACTCAGATCATTTGAACGACTCCATG	9000
9001	Qy	GTCTTAGCGCATTTACACTCCACAGTTACTTCTCCAGTGAGATCAATAGAGTGGCTTCAT	9060
9001	Db	GTCTTAGCGCATTTACACTCCACAGTTACTTCTCCAGTGAGATCAATAGAGTGGCTTCAT	9060
9061	Qy	GCCTCAGGAAACTTGGGGTACCACCTTGGGAACCTGGAGACATCGGCGCAGAGTGTCC	9120
9061	Db	GCCTCAGGAAACTTGGGGTACCACCTTGGGAACCTGGAGACATCGGCGCAGAGTGTCC	9120
9121	Qy	GCCTTAGCTACTTCTCCAGGGGGAGGGCCCACTTCTGCGAGATCACTCTTTAACT	9180
9121	Db	GCCTTAGCTACTTCTCCAGGGGGAGGGCCCACTTCTGCGAGATCACTCTTTAACT	9180
9181	Qy	GGGCAGTAAGGACCAAGCTTAAACTCACTCCAATCCCGCGCGGTCCACGCTGGACTTGT	9240
9181	Db	GGGCAGTAAGGACCAAGCTTAAACTCACTCCAATCCCGCGCGGTCCACGCTGGACTTGT	9240
9241	Qy	CTGGCTAGTTTGGTGGTTCAGCGGGGAGACATATATCAGACCTGTCCTGTCGCC	9300
9241	Db	CTGGCTAGTTTGGTGGTTCAGCGGGGAGACATATATCAGACCTGTCCTGTCGCC	9300
9301	Qy	GACCCCGCTGGTTCCGTTTGTGCTACTCTTCTGTAGGGGTAGGCAATTAACCTGC	9360
9301	Db	GACCCCGCTGGTTTCCGTTTGTGCTACTCTTCTGTAGGGGTAGGCAATTAACCTGC	9360
9361	Qy	TCCCAACCGATGAACGGGAGCTAACCACTCAGGGCTTAAAGCCATTCCTGTTTTT	9420
9361	Db	TCCCAACCGATGAACGGGAGCTAACCACTCAGGGCTTAAAGCCATTCCTGTTTTT	9420
9421	Qy	TT	9480
9421	Db	TT	9480
9481	Qy	TTTTTCCCTTCTTTAATGGTGGCTCCATCTTATAGCCCTAGTCACGGCTAGCTGTGAAAGGT	9540

Db	9481	TTTTTCCTTCTTTAATGATGGTGCATCTTAGCCCTAGTCACGGCTAGCTGTGAAGGT	9541
Qy	9541	CCGTGAGCCGCATGACTGACGACAGAGTGCTGATACTGSCCTCTCTGCAGATCATGT	9595
Db	9541	CCGTGAGCCGCATGACTGACGACAGAGTGCTGATACTGSCCTCTCTGCAGATCATGT	9595
RESULT 2			
ID	AAC86939		
ID	AAC86939	standard; DNA; 9595 BP.	
XX	XX		
AC	AC	AAC86939;	
XX	XX		
DT	02-APR-2001	(first entry)	
XX	XX		
DE	XX	Nucleotide sequence of a hepatitis C virus (HCV) clone genotype 1b.	
XX	XX	Chimeric virus; bovine viral diarrhoea virus; BVDV; hepatitis C virus;	
KW	KW	HCV; vaccine; viral inhibitor; antiviral; ss.	
KW	KW		
XX	XX	Hepatitis C virus.	
OS	XX		
FE	Key	Location/Qualifiers	
FT	CDS	342..9374	
FT	FT	/*tag= a	
XX	XX		
XX	XX	WO200075352-A2.	
PN	XX		
PD	14-DEC-2000.		
XX	XX		
XX	02-JUN-2000;	2000WO-US015527.	
PF	XX		
XX	XX		
PR	04-JUN-1999;	99US-0137817P.	
XX	XX		
PA	(JSSH)	US DEPT HEALTH & HUMAN SERVICES.	
XX	XX		
PI	Nam J, Bukh J, Emerson SU, Purcell RH;		
XX	XX		
DR	WPI; 2001-071081/08.		
DR	P-PSDB; AAB31170.		
XX	XX		
PT	XX	In which the (non-)structural region has been replaced by the (non-)structural	
PT	XX	region of a hepatitis C virus (HCV) genome. The nucleic acids comprising	
PT	XX	the chimeric virus and the chimeric virus are useful for identifying cell	
PT	XX	lines capable of supporting the replication of these chimeric viruses, in	
PT	XX	screening for neutralizing antibodies to HCV of different genotypes, in	
XX	XX	the production of HCV-BVDV virions, for the development of inactivated or	
XX	XX	attenuated vaccines to prevent HCV-BVDV in a mammal, in studying the	
XX	XX	molecular properties of HCV indirectly in vitro, and in identifying	
XX	XX	inhibitors of viral enzyme activity which would be useful as antiviral	
XX	XX	agents. Formulations or compositions comprising the chimeric virions may	
XX	XX	be used to treat or prevent the signs and symptoms of HCV. The present	
XX	XX	sequence represents a HCV clone, which is used to construct chimeric	
XX	XX	nucleic acids of the invention	
PS	Disclosure; Fig 4A-F; 97pp; English.		
XX	XX		
CC	XX	The specification describes a nucleic acid comprising a chimeric virus	
CC	XX	genome, specifically bovine viral diarrhoea virus (BVDV) genome in which	
CC	XX	the (non-)structural region has been replaced by the (non-)structural	
CC	XX	region of a hepatitis C virus (HCV) genome. The nucleic acids comprising	
CC	XX	the chimeric virus and the chimeric virus are useful for identifying cell	
CC	XX	lines capable of supporting the replication of these chimeric viruses, in	
CC	XX	screening for neutralizing antibodies to HCV of different genotypes, in	
CC	XX	the production of HCV-BVDV virions, for the development of inactivated or	
CC	XX	attenuated vaccines to prevent HCV-BVDV in a mammal, in studying the	
CC	XX	molecular properties of HCV indirectly in vitro, and in identifying	
CC	XX	inhibitors of viral enzyme activity which would be useful as antiviral	
CC	XX	agents. Formulations or compositions comprising the chimeric virions may	
CC	XX	be used to treat or prevent the signs and symptoms of HCV. The present	
CC	XX	sequence represents a HCV clone, which is used to construct chimeric	
CC	XX	nucleic acids of the invention	
XX	XX		
SQ	Sequence 9595 BP; 1934 A; 2842 C; 2698 G; 2121 T; 0 U; 0 Other;		
Query Match	100.0%;	Score 9595; D3 4; Length 9595;	
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 9595;	Conservative	0; Mismatches	0; Gaps 0;
QY	1	GCACAGCCCCCTGATGGGGGGGACACTCCACCATGAATCACTCCCTGTGAGAACTACTG	60
Db	1	GCACAGCCCCCTGATGGGGGGGACACTCCACCATGAATCACTCCCTGTGAGAACTACTG	60

Qy	61	TCTTCA	CGCAGAAA	CGCTCTAG	CCATGG	CGTTAG	TATGAG	TGTCGT	CGTG	CAGCCT	CCAGAC	120		
Db	61	TCTTCA	CGCAGAAA	CGCTCTAG	CCATGG	CGTTAG	TATGAG	TGTCGT	CGTG	CAGCCT	CCAGAC	120		
Qy	121	CCCCCT	CCCGGAGAG	CCATAG	TGCTTC	GGGAAC	CGGTGAG	TACAC	CGGAAT	TGCCAG	180			
Db	121	CCCCCT	CCCGGAGAG	CCATAG	TGCTTC	GGGAAC	CGGTGAG	TACAC	CGGAAT	TGCCAG	180			
Qy	181	GACGAC	CGGGTCC	TTTTCT	TGTGAT	CAAC	CGCTCA	ATGCT	GGAGAT	TTGGCG	TGCCCCC	240		
Db	181	GACGAC	CGGGTCC	TTTTCT	TGTGAT	CAAC	CGCTCA	ATGCT	GGAGAT	TTGGCG	TGCCCCC	240		
Qy	241	CGGAGACT	GCTPAG	CGAGTAG	TGTTGG	TTCG	CGAAAG	CCCTP	TGTGTA	CTGCCT	GATAGG	300		
Db	241	CGGAGACT	GCTPAG	CGAGTAG	TGTTGG	TTCG	CGAAAG	CCCTP	TGTGTA	CTGCCT	GATAGG	300		
Qy	301	GTGCTT	GGCAGT	CCCCCG	GGGTCT	CTGTAG	ACCG	TGACCAT	GAGCAG	AACTCT	TAAC	360		
Db	301	GTGCTT	GGCAGT	CCCCCG	GGGTCT	CTGTAG	ACCG	TGACCAT	GAGCAG	AACTCT	TAAC	360		
Qy	361	CTCAAG	AAAAAC	CAACG	TAAACA	CAAC	CGCG	CCCA	CAGGAC	GTCAAG	TTC	CGGCG	420	
Db	361	CTCAAG	AAAAAC	CAACG	TAAACA	CAAC	CGCG	CCCA	CAGGAC	GTCAAG	TTC	CGGCG	420	
Qy	421	GTGCT	CAGATCG	TTGGT	GGAGTT	TTACT	CTGTT	CCCG	CAGGG	CCCCC	AGTTGG	TGCTGC	480	
Db	421	GTGCT	CAGATCG	TTGGT	GGAGTT	TTACT	CTGTT	CCCG	CAGGG	CCCCC	AGTTGG	TGCTGC	480	
Qy	481	CGCGCA	CTAG	GAAGC	TTCG	AGCG	GTGC	GAAC	CTCG	TGG	AAAGG	CAAC	CTAT	540
Db	481	CGCGCA	CTAG	GAAGC	TTCG	AGCG	GTGC	GAAC	CTCG	TGG	AAAGG	CAAC	CTAT	540
Qy	541	AGGCT	CGCG	ACCCG	AGGCG	CTCG	GGCT	CAGCC	CGGT	CAC	CTT	GGCC	CT	600
Db	541	AGGCT	CGCG	ACCCG	AGGCG	CTCG	GGCT	CAGCC	CGGT	CAC	CTT	GGCC	CT	600
Qy	601	GCAAT	GAGG	CGCT	GGGT	TGGC	AGGAT	GGCT	CTCT	CA	CCCG	GGGCT	CCCG	660
Db	601	GCAAT	GAGG	CGCT	GGGT	TGGC	AGGAT	GGCT	CTCT	CA	CCCG	GGGCT	CCCG	660
Qy	661	GGGGCCC	AC	CGAC	CCCCCG	GGTAG	TGCG	TAACT	TGG	TAA	AGT	TCAT	CGAT	720
Db	661	GGGGCCC	AC	CGAC	CCCCCG	GGTAG	TGCG	TAACT	TGG	TAA	AGT	TCAT	CGAT	720
Qy	721	CAT	GCGG	CTCG	CGAT	CTCAT	GGGT	TACAT	TCG	CT	CGG	CGCC	CCCT	780
Db	721	CAT	GCGG	CTCG	CGAT	CTCAT	GGGT	TACAT	TCG	CT	CGG	CGCC	CCCT	780
Qy	781	CTGCC	AGG	CGCT	TGG	CA	CGGT	GTCC	GGGT	CT	TGG	AGG	CGGT	840
Db	781	CTGCC	AGG	CGCT	TGG	CA	CGGT	GTCC	GGGT	CT	TGG	AGG	CGGT	840
Qy	841	GGAACT	TC	CGG	TGCT	CTT	CTAT	TC	CTT	TG	CT	TC	TG	900

Qy	1141	GGACGGCTGCTTTCTGCTCCGCTAAGTACGHTGGGGATCTCTGGGATCTATTTTTCTCTCG	1200
Db	1141	GGACGGGTGCTTTTCTGCTCCGCTATGATACGTGGGGATCTCTGGGATCTATTTTTCTCTCG	1200
Qy	1201	TCTCCCGCTGTTCACCTTTCGCGCTCGCGGCATGAGACGTGCAGACTCGCAACTGCT	1260
Db	1201	TCTCCCGAGCTTTCACCTTTCGCGCTCGCGGCATGAGACGTGCAGACTCGCAACTGCT	1260
Qy	1261	CAATCTATCCCGGCATGTATCAGTCAACGCATGGCTTGGGATATGATGATCAACTGGT	1320
Db	1261	CAATCTATCCCGGCATGTATCAGTCAACGCATGGCTTGGGATATGATGATCAACTGGT	1320
Qy	1321	CACCTACAAAGCCCTTATGTGTGTCGCGATGCCGATCCCAACAGCTGTCTGTGACAC	1380
Db	1321	CACCTACAAAGCCCTTATGTGTGTCGCGATGGCTTCCGATCCCAACAGCTGTCTGTGACAC	1380
Qy	1381	TGGTGGCGGGGCCACATGGGGAGTCTGGGGGCCCTTGCCTACTATTCATGTTAGGGA	1440
Db	1381	TGGTGGCGGGGCCACATGGGGAGTCTGGGGGCCCTTGCCTACTATTCATGTTAGGGA	1440
Qy	1441	ACTGGGCTTAACTTCTGATTTGTGGCGTACTCTTTTGGCGGCGTTGACGGGGAGACCCACA	1500
Db	1441	ACTGGGCTTAACTTCTGATTTGTGGCGTACTCTTTTGGCGGCGTGAACGGGGAGACCCACA	1500
Qy	1501	CGACGGGAGGGTGCGCGGCCACACCACTCCGGTTCAAGTCCCTTTTCTCACTCTGGG	1560
Db	1501	CGACGGGAGGGTGCGCGGCCACACCACTCCGGGTTCAAGTCCCTTTTCTCATCTGGG	1560
Qy	1561	CGTCTCAGAAATTCAGCTTGTGAATACCAACGCGACGTGGCACATCAACAGACTGCC	1620
Db	1561	CGTCTCAGAAATTCAGCTTGTGAATACCAACGCGACGTGGCACATCAACAGGACTGCC	1620
Qy	1621	TAAATTCGAATGACTCCCTCCAAACTGGGTCTTTTGGCGGCTCTTTTACGCACAAGT	1680
Db	1621	TAAATTCGAATGACTCCCTCCAAACTGGGTCTTTTGGCGGCTGTTTACGCACAAGT	1680
Qy	1681	TCAACTCGTCCGGTGCCCGGAGCGCACAGCTCGCGCCCACTGACTGTTTGGGCC	1740
Db	1681	TCAACTCGTCCGGTGCCCGGAGCGCATGCGACGTCCGCGCCCACTGACTGTTTGGGCC	1740
Qy	1741	AGGGTGGGGCCCATCACCTATCTAAGCTTAAACACTGGATCGGATCGGACCTTATTTGCT	1800
Db	1741	AGGGGTGGGGCCCATCACCTATCTAAGCTTAAACACTGGATCGGATCGGACCTTATTTGCT	1800
Qy	1801	GGCATTAACGGCCCTCGACCGGTGGTGGTACCGCGGTGCGAGCTGTGGTCCAGTGT	1860
Db	1801	GGCATTAACGGCCCTCGACCGGTGGTGGTACCGCGGTGCGAGCTGTGGTCCAGTGT	1860
Qy	1861	ATTGTTTACCCCCAAGCCCTTTGTTGGTGGGACCAACCGATCGTTCGGTGTCCCTACGT	1920
Db	1861	ATTGTTTACCCCCAAGCCCTTTGTTGGTGGGACCAACCGATCGTTCGGTGTCCCTACGT	1920
Qy	1921	ATAGCTGGGGGAGAAATGACACAGCTGATGCTCTTCAACACACGCTCCGGCCACAAG	1980
Db	1921	ATAGCTGGGGGAGAAATGACACAGCTGATGCTCTTCAACACACGCTCCGGCCACAAG	1980
Qy	1981	GCAACTGTTTCGGCTGTACATGATGAATAGTACTTGGGTTTCACTAAGACGTGGGAGTTC	2040
Db	1981	GCAACTGTTTCGGCTGTACATGATGAATAGTACTTGGGTTTCACTAAGACGTGGGAGTTC	2040
Qy	2041	CCCCGTGTAAATCGGGGGGTTCGGTAAACCGCACCTTGTATCTGCCCAACGACTGCTCC	2100
Db	2041	CCCCGTGTAAATCGGGGGGTTCGGTAAACCGCACCTTGTATCTGCCCAACGACTGCTCC	2100
Qy	2101	GGAGGACCCCGAGGCTACTTACACAAATGTGGCTCGGGGCCCTGTTGACACCTTAGT	2160
Db	2101	GGAGGACCCCGAGGCTACTTACACAAATGTGGCTCGGGGCCCTGTTGACACCTTAGT	2160
Qy	2161	GCCTAGTAGACTACCCATACAGGCTTTGGCACTACCCCTGCACCTCTCAATTTTTCCTCT	2220
Db	2161	GCCTAGTAGACTACCCATACAGGCTTTGGCACTACCCCTGCACCTCTCAATTTTTCCTCT	2220
Qy	2221	TTAAGGTTTAGATGATGTGTGGGGGGGTGGAGCACAGGCTCAATGCGCGCATGCAATTGGA	2280

Db 2221 TTAAGCTTAGATGTATGTGGGGGGGTGGAGACAGGCTCAATGCGCATGCAATTGGA 2280
QY 2281 CTGAGGAGAGCCCTGTAACTTGGAGGACAGGATAGTCAAGACTCAGCCCGCTCTGTC 2340
Db 2281 CTGAGGAGAGCCCTGTAACTTGGAGGACAGGATAGTCAAGACTCAGCCCGCTCTGTC 2340
QY 2341 TGTCTCAACAGAGAGTGGACATAGTCCCTGTGCTTCCACACCTTACCGCTTTATCCA 2400
Db 2341 TGTCTCAACAGAGTGGACATAGTCCCTGTGCTTCCACACCTTACCGCTTTATCCA 2400
QY 2401 CTGGTTTGAATCCATCCAGAAATCCGTGAGCGTGGAATACCTGTAGGTAGGGT 2460
Db 2401 CTGGTTTGAATCCATCCAGAAATCCGTGAGCGTGGAATACCTGTAGGTAGGGT 2460
QY 2461 CAGCGTTTGTCTCTTGTGAATCAATGAGGATACATCCGTGCTTTTCTCTCTG 2520
Db 2461 CAGCGTTTGTCTCTTGTGAATCAATGAGGATACATCCGTGCTTTTCTCTCTG 2520
QY 2521 CAGACGCGCGGTGTGCTGCTTGTGATGATGCTGTATAGCCCAAGCTGAGGCCG 2580
Db 2521 CAGACGCGCGGTGTGCTGCTTGTGATGATGCTGTATAGCCCAAGCTGAGGCCG 2580
QY 2581 CTTTGAAGAACTTGTGTCTCTCAATGCGCGCTCCGTGGCGGAGCGCATGTATCTCT 2640
Db 2581 CTTTGAAGAACTTGTGTCTCTCAATGCGCGCTCCGTGGCGGAGCGCATGTATCTCT 2640
QY 2641 CTTTCTTGTCTCTTGTGCGCGCTGTGATTAAGGACAGGCTGCTTCTGCGCGG 2700
Db 2641 CTTTCTTGTCTCTTGTGCGCGCTGTGATTAAGGACAGGCTGCTTCTGCGCGG 2700
QY 2701 CGTATGCTTTTATGCGGTATGCGCGCTGCTCTGCTCTCTACTGCGGTTACACACGAG 2760
Db 2701 CGTATGCTTTTATGCGGTATGCGCGCTGCTCTGCTCTCTACTGCGGTTACACACGAG 2760
QY 2761 CTTTACCCCTTGGACCGGAGATGCTGATCGTGGCGGGTGGGTTCTTGTAGTCTGG 2820
Db 2761 CTTTACCCCTTGGACCGGAGATGCTGATCGTGGCGGGTGGGTTCTTGTAGTCTGG 2820
QY 2821 TATTCTTGTACTTGTACCATATCAAAAGTGTCTCACTAGGCTCATATGTTGTTAC 2880
Db 2821 TATTCTTGTACTTGTACCATATCAAAAGTGTCTCACTAGGCTCATATGTTGTTAC 2880
QY 2881 AATACTTTATACAGAGCGGAGCGGACATGCAAGTGTGGTCCCGCCCTCAACGTTT 2940
Db 2881 AATACTTTATACAGAGCGGAGCGGACATGCAAGTGTGGTCCCGCCCTCAACGTTT 2940
QY 2941 GGGAGGCGCGATGCCATCATCTCTCAAGTGTGGGTTCAACAGATTAATTTTG 3000
Db 2941 GGGAGGCGCGATGCCATCATCTCTCAAGTGTGGGTTCAACAGATTAATTTTG 3000
QY 3001 ACATCAACCAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060
Db 3001 ACATCAACCAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060
QY 3061 CGAGATGCGGTACTTGTGCGGCTCAAGGCTCATCTGTGATGATGATGATGATGATGAT 3120
Db 3061 CGAGATGCGGTACTTGTGCGGCTCAAGGCTCATCTGTGATGATGATGATGATGATGAT 3120
QY 3121 AAGTCCCGGGGTCAATATGTCCTCAATGCTTCAATGCTGCTGCTGCTGCTGCTGCTGCT 3180
Db 3121 AAGTCCCGGGGTCAATATGTCCTCAATGCTTCAATGCTGCTGCTGCTGCTGCTGCTGCT 3180
QY 3181 CGTACGCTTATAACCATCTTACCCACTGCGGACTGGGCCACGCGGGCTACGAGACC 3240
Db 3181 CGTACGCTTATAACCATCTTACCCACTGCGGACTGGGCCACGCGGGCTACGAGACC 3240
QY 3241 TTGCGGTGCGGTAGAGCCCGTCTTCTTCCGCGATGAGACCAAGGTCATCACTGGG 3300
Db 3241 TTGCGGTGCGGTAGAGCCCGTCTTCTTCCGCGATGAGACCAAGGTCATCACTGGG 3300
QY 3301 GAGCAGACACCGCTGCTGGTGGGACATCATCTTGGTCTTACCGTCTCGCCCGAGGG 3360

Db 3301 GAGCAGACACCGCTGCTGGTGGGACATCATCTTGGTCTACCCGCTCTCGCCCGAGGG 3360
QY 3361 GGAAGAGATATTTTGGGACCGCTGATAGTCTCGAAGGCAAGGTCGGCACTCCTTGG 3420
Db 3361 GGAAGAGATATTTTGGGACCGCTGATAGTCTCGAAGGCAAGGTCGGCACTCCTTGG 3420
QY 3421 CGCCCATCAACGCGCTACTCCCAACAAACGCGGGGCTACTTGGTTGCATCATCACTAGCC 3480
Db 3421 CGCCCATCAACGCGCTACTCCCAACAAACGCGGGGCTACTTGGTTGCATCATCACTAGCC 3480
QY 3481 TCACAGGCGGAGCAAGAACAGGTGCAAGGGGAGGTTCAAGTGTGTCTTACCGCAACAC 3540
Db 3481 TCACAGGCGGAGCAAGAACAGGTGCAAGGGGAGGTTCAAGTGTGTCTTACCGCAACAC 3540
QY 3541 AATCTTTCTGCGGACTGATCAACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
Db 3541 AATCTTTCTGCGGACTGATCAACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
QY 3601 CGAAGACCTTAGCCGCTCCAAAGGTCCAAATCACCCAAATGTACCAATGTAGACCTGG 3660
Db 3601 CGAAGACCTTAGCCGCTCCAAAGGTCCAAATCACCCAAATGTACCAATGTAGACCTGG 3660
QY 3661 ACTCGTGGCTGGCAGGCGCCCGCGGGCGCTCCATGACACCATGACGTGTGCA 3720
Db 3661 ACTCGTGGCTGGCAGGCGCCCGCGGGCGCTCCATGACACCATGACGTGTGCA 3720
QY 3721 GCTCGGACCTTTACTTGTTCAGGACATGCTGATGCTCAATCCGGTGGCGGAGGCG 3780
Db 3721 GCTCGGACCTTTACTTGTTCAGGACATGCTGATGCTCAATCCGGTGGCGGAGGCG 3780
QY 3781 ACAGAGGGGAAHTACTCTTCCCGCAGGCGCGCTCTCTTACCTGAAAGGCTCTCGGCTG 3840
Db 3781 ACAGAGGGGAAHTACTCTTCCCGCAGGCGCGCTCTCTTACCTGAAAGGCTCTCGGCTG 3840
QY 3841 GTCCATGCTTTCCTTGGGCGAGCTGCTGGGCGCTCTTCCGGGCTGCTGTGCAACC 3900
Db 3841 GTCCATGCTTTCCTTGGGCGAGCTGCTGGGCGCTCTTCCGGGCTGCTGTGCAACC 3900
QY 3901 GGGGGGTGCGAAGGCGGTGGACTTACACCGCTGAGTCTATGAAACTTACCATGCGGT 3960
Db 3901 GGGGGGTGCGAAGGCGGTGGACTTACACCGCTGAGTCTATGAAACTTACCATGCGGT 3960
QY 3961 CTCCGCTCTTCAACAGCAACTCAACCCCGCGCTGTACCGAGACATTCOAAGTGGAC 4020
Db 3961 CTCCGCTCTTCAACAGCAACTCAACCCCGCGCTGTACCGAGACATTCOAAGTGGAC 4020
QY 4021 ATCTGCACTCTTCTGCGAGCGGACAGACCAAAAGTGGCGCTGCTGCTGCTGCTGCTGCT 4080
Db 4021 ATCTGCACTCTTCTGCGAGCGGACAGACCAAAAGTGGCGCTGCTGCTGCTGCTGCTGCT 4080
QY 4081 AAGGGTACAAAGTGTCTGCTCTGAAACCGCTGCGGCTTACCGCTTACCGCTTGGGGGT 4140
Db 4081 AAGGGTACAAAGTGTCTGCTCTGAAACCGCTGCGGCTTACCGCTTACCGCTTGGGGGT 4140
QY 4141 ATATGTCGAAGGACACAGGTATGACCTTACATGAGAACTGGGGTAAAGGACCATTACCA 4200
Db 4141 ATATGTCGAAGGACACAGGTATGACCTTACATGAGAACTGGGGTAAAGGACCATTACCA 4200
QY 4201 CGGGCGGCTTCAATGAGTACTCCACTATGCAAGTCTTCCCGGAGCGGTGGCTTCTG 4260
Db 4201 CGGGCGGCTTCAATGAGTACTCCACTATGCAAGTCTTCCCGGAGCGGTGGCTTCTG 4260
QY 4261 GGGGCGCTTATGACATCATATATGATGATGCTGCTCACTCACTGACTACCATCT 4320
Db 4261 GGGGCGCTTATGACATCATATATGATGATGCTGCTCACTCACTGACTACCATCT 4320
QY 4321 TGGGATCGGACAGTCTTGGACCAAGCGGAGACGCTGGAGCGGCTCTGCTGCTGCTGCTG 4380
Db 4321 TGGGATCGGACAGTCTTGGACCAAGCGGAGACGCTGGAGCGGCTCTGCTGCTGCTGCTG 4380
QY 4381 CCACCGCTACCTCCGGGATCGGTTACCGTGGCCACACCCCAATATCCAGGAATAGGCC 4440
Db 4381 CCACCGCTACCTCCGGGATCGGTTACCGTGGCCACACCCCAATATCCAGGAATAGGCC 4440

Qy	4441	TGTTCCAAACATGAGAGATCCCTTCTATGCAAGCCATCCCCATTGAGGCCATCAAGG	4500
Db	4441	TGTTCCAAACATGAGAGATCCCTTCTATGCAAGCCATCCCCATTGAGGCCATCAAGG	4500
Qy	4501	GGGGAGGCATCTCATTTCTGCCATTCCAAAGAAATGACGAGCTGCCCGCAAGC	4560
Db	4501	GGGGAGGCATCTCATTTCTGCCATTCCAAAGAAATGACGAGCTGCCCGCAAGC	4560
Qy	4561	TGACAGGCCTCGACTGAACCTGTAGCATATTACCGGGGCTTGATGTGTCGCTCATAC	4620
Db	4561	TGACAGGCCTCGACTGAACCTGTAGCATATTACCGGGGCTTGATGTGTCGCTCATAC	4620
Qy	4621	CGCCTATCGGAGAGCTGCTGCTGTGCGCAACAGACGCTTAATGACGCGTTTCAACCGCGC	4680
Db	4621	CGCCTATCGGAGAGCTGCTGCTGTGCGCAACAGACGCTTAATGACGCGTTTCAACCGCGC	4680
Qy	4681	ATTTTGACTCAGTGATCGACTGCAATACATGTGTACCCAGACAGTTCGACTTCAGCTTGG	4740
Db	4681	ATTTTGACTCAGTGATCGACTGCAATACATGTGTACCCAGACAGTTCGACTTCAGCTTGG	4740
Qy	4741	ATCCCACTTCACCATTTAGACGACGACCTGCCCCCAAGACGCGGTGTGCGCTTCGCAAC	4800
Db	4741	ATCCCACTTCACCATTTAGACGACGACCTGCCCCCAAGACGCGGTGTGCGCTTCGCAAC	4800
Qy	4801	GGCGAGGTAGAACTGGCAGGGGTAGGATGCGCATCTACAGCTTTGTGACTCCAGAGAAC	4860
Db	4801	GGCGAGGTAGAACTGGCAGGGGTAGGATGCGCATCTACAGCTTTGTGACTCCAGAGAAC	4860
Qy	4861	GGCCCTCGGGCATGTTCTGATTTCTCGTCTGTGTGAGTGTATGACGCGGGCTGCTT	4920
Db	4861	GGCCCTCGGGCATGTTCTGATTTCTCGTCTGTGTGAGTGTATGACGCGGGCTGCTT	4920
Qy	4921	GGTATGAGCTCACGCCCTCGAGACCTCGGTAGCTTTCGCGGGCTTAACTAAATACACAG	4980
Db	4921	GGTATGAGCTCACGCCCTCGAGACCTCGGTAGCTTTCGCGGGCTTAACTAAATACACAG	4980
Qy	4981	GGTTGCCCGTCTGCCAGGACCATCTGGAGTTCTCGGAGAGCGTCTTACAGGCTCACCC	5040
Db	4981	GGTTGCCCGTCTGCCAGGACCATCTGGAGTTCTCGGAGAGCGTCTTACAGGCTCACCC	5040
Qy	5041	ACATAGATGCCACTTCTCTGCCAGACTAAACAGGCAGAGAGCAACTTTCCTTCACTGCG	5100
Db	5041	ACATAGATGCCACTTCTCTGCCAGACTAAACAGGCAGAGAGCAACTTTCCTTCACTGCG	5100
Qy	5101	TGGCATATCAAGCTACAGTGTGCCCGAGGGCTCAAGCTCCACCTTCATCTGGGACCAAA	5160
Db	5101	TGGCATATCAAGCTACAGTGTGCCCGAGGGCTCAAGCTCCACCTTCATCTGGGACCAAA	5160
Qy	5161	TGTGGAGTGTCTCATAGGCTGAACCTACATGCGAGGGCCAAACCCCTGCTGTATA	5220
Db	5161	TGTGGAGTGTCTCATAGGCTGAACCTACATGCGAGGGCCAAACCCCTGCTGTATA	5220
Qy	5221	GGCTTAGAGCGCTCCAAATAGAGTCACTTCACACACCCCAATTAATACATCATGG	5280
Db	5221	GGCTTAGAGCGCTCCAAATAGAGTCACTTCACACACCCCAATTAATACATCATGG	5280
Qy	5281	CATGATGTTCGGCTGACCTGGAGTGTCTACATAGCAGCTGGTGTGATTTGAGCGGAGTCC	5340
Db	5281	CATGATGTTCGGCTGACCTGGAGTGTCTACATAGCAGCTGGTGTGATTTGAGCGGAGTCC	5340
Qy	5341	TTGCGAGTTTGGCGGATATGCTGCTGACGACAGGCACTGTGCTTGTGGCGAGGATCA	5400
Db	5341	TTGCGAGTTTGGCGGATATGCTGCTGACGACAGGCACTGTGCTTGTGGCGAGGATCA	5400
Qy	5401	TCTTGTCCGGAGCCAGCTGTCTGTTCCCGACAGGAGTCTCTACAGAGTTTGATG	5460
Db	5401	TCTTGTCCGGAGCCAGCTGTCTGTTCCCGACAGGAGTCTCTACAGAGTTTGATG	5460
Qy	5461	AGATGGAGAGTGTGCTTCAAACTTCTTACATCGAGAGGGAAATGACGCTGCCGAGC	5520
Db	5461	AGATGGAGAGTGTGCTTCAAACTTCTTACATCGAGAGGGAAATGACGCTGCCGAGC	5520

Qy	5521	AATTCAAGCAAAAGCGCGCTCGGCTTGTTGTCGAAACGGGCCACAAAGCAAGCGGAGGCTGCTG	5580
Db	5521	AATTCAAGCAAAAGCGCGCTCGGCTTGTTGTCGAAACGGGCCACAAAGCAAGCGGAGGCTGCTG	5590
Qy	5581	CTCCCGTGGTGGAGTCCAAAGTGGCGAGCCCTTCAGACCTTCGGCGAAACACATGTGGA	5640
Db	5581	CTCCCGTGGTGGAGTCCAAAGTGGCGAGCCCTTCAGACCTTCGGCGAAACACATGTGGA	5640
Qy	5641	ATTTTCATCAGCGGAATACAGTACCTAGCAGGCTTATCCACTCTGCTTGGAAACCCCGCGA	5700
Db	5641	ATTTTCATCAGCGGAATACAGTACCTAGCAGGCTTATCCACTCTGCTTGGAAACCCCGCGA	5700
Qy	5701	TAGCATCATTTATGGCATTTTACAGCTTCTATCATAGCCCGCTCAACCCAAAAACACC	5760
Db	5701	TAGCATCATTTATGGCATTTTACAGCTTCTATCATAGCCCGCTCAACCCAAAAACACC	5760
Qy	5761	TCCTGTTTAAATCTCTGGGGGATGGGTGGCTGCCAACTCGCTCTCCACAGCGCTGCGT	5820
Db	5761	TCCTGTTTAAATCTTGGGGGATGGGTGGCTGCCAACTCGCTCTCCACAGCGCTGCGT	5820
Qy	5821	CAGCTTTCTGGGCGCGCGGATCGCCGAGCGCTGTGGACATAGGCTTGGGAAAG	5880
Db	5821	CAGCTTTCTGGGCGCGCGGATCGCCGAGCGCTGTGGACATAGGCTTGGGAAAG	5880
Qy	5881	TGCTCTGGGACATCTTGGCGGGCTATGGGCGAGGGTAGCGCGCACTCGTGGCCTTTA	5940
Db	5881	TGCTCTGGGACATCTTGGCGGGCTATGGGCGAGGGTAGCGCGCACTCGTGGCCTTTA	5940
Qy	5941	AGGTCATGAGCGGAGTGCCCTCCACAGAGACCTGTCACACTTACTCCCTGCCATCC	6000
Db	5941	AGGTCATGAGCGGAGTGCCCTCCACAGAGACCTGTCACACTTACTCCCTGCCATCC	6000
Qy	6001	TCTCTCTGTGGCTTGGTCTCGGGGTCTGTGCGCAGCAATACTCGCTCGGCACGTGG	6060
Db	6001	TCTCTCTGTGGCTTGGTCTCGGGGTCTGTGCGCAGCAATACTCGCTCGGCACGTGG	6060
Qy	6061	GCCCGGAGAGGGGCTGTCACTGCGATGAACCCGGCTGATAGCGTTTCGTCGCGGGTA	6120
Db	6061	GCCCGGAGAGGGGCTGTCACTGCGATGAACCCGGCTGATAGCGTTTCGTCGCGGGTA	6120
Qy	6121	ACCAAGTCTCCCTACGACATATGTGCTCAGAGCGACGCTGACGACAGCTGTCACTCAGA	6180
Db	6121	ACCAAGTCTCCCTACGACATATGTGCTCAGAGCGACGCTGACGACAGCTGTCACTCAGA	6180
Qy	6181	TCTCTCTAGCCTTACCATCACTCAACTGCTGTAAGCGGCTCCACCAAGTGATTAATGAGG	6240
Db	6181	TCTCTCTAGCCTTACCATCACTCAACTGCTGTAAGCGGCTCCACCAAGTGATTAATGAGG	6240
Qy	6241	ACTGCTTAGCCCATGTCCGGCTCGTGGCTAAGGATGTTTGGATTTGGATATGCACGG	6300
Db	6241	ACTGCTTAGCCCATGTCCGGCTCGTGGCTAAGGATGTTTGGATTTGGATATGCACGG	6300
Qy	6301	TGTTGACTGACTTCAGACCTGCTCCAGTCAAACTCTGCGCGGTTACCGGAGTCC	6360
Db	6301	TGTTGACTGACTTCAGACCTGCTCCAGTCAAACTCTGCGCGGTTACCGGAGTCC	6360
Qy	6361	CTTTCTCTGTATGCCAAACCGGGTAAAGGAGTCTTGGCGGGGACGGCATCATGCAAAA	6420
Db	6361	CTTTCTCTGTATGCCAAACCGGGTAAAGGAGTCTTGGCGGGGACGGCATCATGCAAAA	6420
Qy	6421	CCAACCTGCCCATGCGGAGCACAGATGCCCGGACATGTCAAAAACGGTTCCATAGAGATCG	6480
Db	6421	CCAACCTGCCCATGCGGAGCACAGATGCCCGGACATGTCAAAAACGGTTCCATAGAGATCG	6480
Qy	6481	TAGGGCTAGAACTGTGAGCAACAGTGGCACGGAGCTTCCCATCAACGATACACCA	6540
Db	6481	TAGGGCTAGAACTGTGAGCAACAGTGGCACGGAGCTTCCCATCAACGATACACCA	6540
Qy	6541	CGGACCTTGCACACCTCCCGCGGCCCAACTATTTCAGGGGCGTATGCGGGTGGCTG	6600
Db	6541	CGGACCTTGCACACCTCCCGCGGCCCAACTATTTCAGGGGCGCTATGCGGGTGGCTG	6600
Qy	6601	CTGAGGAGTACGTGGAGTTTACGCTGTCTGGGGATTTCCACTACGTGACGGGCATCAACCA	6660

Db 5601 CTGAGGAGTACGTGGAGGTTACGCGTGTGGGGATTTCACACTACGTGACGGGCAATGACCA 6660
Qy 5661 CTGACACGTTAAAGTGCCTATGCCAGTTCGGGCCCGCCGAAATCTTCAACGGAGTGGATG 6720
Db 5661 CTGACACGTTAAAGTGCCTATGCCAGTTCGGGCCCGCCGAAATCTTCAACGGAGTGGATG 6720
Qy 6721 GAGTGGGGTTGCACAGGTACGCTCCGGCGTGCMAACCTCTTCTACCGGAGGAGGTCAAGT 6780
Db 6721 GAGTGGGGTTGCACAGGTACGCTCCGGCGTGCMAACCTCTTCTACCGGAGGAGGTCAAGT 6780
Qy 6781 TCCAGGTTCGGGCTCAACCAATACCTGGTTCGGGTTCGAGCTCCCATGAGAGCCCGAACCGG 6840
Db 6781 TCCAGGTTCGGGCTCAACCAATACCTGGTTCGGGTTCGAGCTCCCATGAGAGCCCGAACCGG 6840
Qy 6841 ACCTAAACAGTGTCTACTCCATGCTACCGATCCCTCCCAATACAGCAGAGACAGGCTA 6900
Db 6841 ACCTAAACAGTGTCTACTCCATGCTACCGATCCCTCCCAATACAGCAGAGAGGCTA 6900
Qy 6901 AGCGTAGGCTGGGTAGAGGTCTCCCCCTCTTTAGCCAGCTCATAGCTAGCCAGTTGT 6960
Db 6901 AGCGTAGGCTGGGTAGAGGTCTCCCCCTCTTTAGCCAGCTCATAGCTAGCCAGTTGT 6960
Qy 6961 CTGGCCTCTTTGAAGGCGACATGACTACCAACCAATGATCTCCCGGAGCTGACTCA 7020
Db 6961 CTGGCCTCTTTGAAGGCGACATGACTACCAACCAATGATCTCCCGGAGCTGACTCA 7020
Qy 7021 TCGAGGCCAACCTCTTTGTGGCGGACAGAGATGGCGGAAACATCACTCGCGTGGAGTCA 7080
Db 7021 TCGAGGCCAACCTCTTTGTGGCGGACAGAGATGGCGGAAACATCACTCGCGTGGAGTCA 7080
Qy 7081 AGAATAGGTAGTAAATCTCGACTCTTTTCGAAACCGCTTCACCGGAGGGGATGAGAGG 7140
Db 7081 AGAATAGGTAGTAAATCTCGACTCTTTTCGAAACCGCTTCACCGGAGGGGATGAGAGG 7140
Qy 7141 AGATATCCGTCGGCGGAGATCTCGGAAATCCAGGAGTTCCCTGAGCGTTCGCCA 7200
Db 7141 AGATATCCGTCGGCGGAGATCTCGGAAATCCAGGAGTTCCCTGAGCGTTCGCCA 7200
Qy 7201 TATGGSCACGCCCGGACTACAATCTCTCACTGTAGAGTCTCGGAAAGACCCGGACTACG 7260
Db 7201 TATGGSCACGCCCGGACTACAATCTCTCACTGTAGAGTCTCGGAAAGACCCGGACTACG 7260
Qy 7261 TCCCTCGGTGTGACAGGATGCCATTCGCACTCAAGGTCCTCCAAATACACCTC 7320
Db 7261 TCCCTCGGTGTGACAGGATGCCATTCGCACTCAAGGTCCTCCAAATACACCTC 7320
Qy 7321 CACGGAGAAAGAGGAGGTGTCTGACAGATCCAAATGTCCTCTGCTTGGCGGAGC 7380
Db 7321 CACGGAGAAAGAGGAGGTGTCTGACAGATCCAAATGTCCTCTGCTTGGCGGAGC 7380
Qy 7381 TCGCCACTAAGACTTCGGTAGTTCGGATCGTCGGCCGTTGATAGCGGACCGGACCG 7440
Db 7381 TCGCCACTAAGACTTCGGTAGTTCGGATCGTCGGCCGTTGATAGCGGACCGGACCG 7440
Qy 7441 CCTTCTGACCTCGCTCCGACGAGGTGACAAAGATCCGAGTTCGATCTCCT 7500
Db 7441 CCTTCTGACCTCGCTCCGACGAGGTGACAAAGATCCGAGTTCGATCTCCT 7500
Qy 7501 CCATGCCCCCTTGAAGGGGAGCGGGGACCCCGATCTCAGCGAGGTCCTTGGTCTA 7560
Db 7501 CCATGCCCCCTTGAAGGGGAGCGGGGACCCCGATCTCAGCGAGGTCCTTGGTCTA 7560
Qy 7561 CCGTAGTAGGAGGTAGTAGGATGCTGCTGCTGCTCAATGTCCTATACGTGGACAG 7620
Db 7561 CCGTAGTAGGAGGTAGTAGGATGCTGCTGCTGCTCAATGTCCTATACGTGGACAG 7620
Qy 7621 GCGCCCTGATCAGCCATGCGCTCGGAGGAAAGTAAAGTGCCTCAACCCCTTAGCA 7680
Db 7621 GCGCCCTGATCAGCCATGCGCTCGGAGGAAAGTAAAGTGCCTCAACCCCTTAGCA 7680
Qy 7681 ACTCTTTGCTGCTGCTACCAACATGCTTACGCCACACATCCCGCAGCGGAGCTCC 7740

Db 7681 ACTCTTTGCTGCTGCTACCAACATGCTTACGCCACACATCCCGCAGCGAACCTCC 7740
Qy 7741 GGCAGAGAGAGTCACTTTGACAGATTCGAAGTTCCTGATGATCATTACCGGAGCTAC 7800
Db 7741 GGCAGAGAGAGTCACTTTGACAGATTCGAAGTTCCTGATGATCATTACCGGAGCTAC 7800
Qy 7801 TCAAGGAGATGAAGGGAAGCGGTCCACAGTTTAAGCTTAAGCTTCTATCTATAGGAGG 7860
Db 7801 TCAAGGAGATGAAGGGAAGCGGTCCACAGTTTAAGCTTAAGCTTCTATCTATAGGAGG 7860
Qy 7861 CTTGMAAGTGAAGCGCCCAATTCGGGCAATCAAAATTTGGCTATGGGCAAGGAGG 7920
Db 7861 CTTGMAAGTGAAGCGCCCAATTCGGGCAATCAAAATTTGGCTATGGGCAAGGAGG 7920
Qy 7921 TCGGAACTATCCAGCAGGCGCTTAACCAATCCGTCCTGCTGGGAGGACTTGTGG 7980
Db 7921 TCGGAACTATCCAGCAGGCGCTTAACCAATCCGTCCTGCTGGGAGGACTTGTGG 7980
Qy 7981 AAGACACTGAACACCAATTTGACACCAATCATGATGCAAAAGTGAAGTTTCTCGTCC 8040
Db 7981 AAGACACTGAACACCAATTTGACACCAATCATGATGCAAAAGTGAAGTTTCTCGTCC 8040
Qy 8041 AACCGAGAGAGGAGCGCCGACAGCTGCGCTTATCGTATCCAGACCTGGAGTTC 8100
Db 8041 AACCGAGAGAGGAGCGCCGACAGCTGCGCTTATCGTATCCAGACCTGGAGTTC 8100
Qy 8101 GTGTATGCGAGAGATGGGCCCTTTACGAGCTGTCTCCACCTTCTCAGGCGCTGATGG 8160
Db 8101 GTGTATGCGAGAGATGGGCCCTTTACGAGCTGTCTCCACCTTCTCAGGCGCTGATGG 8160
Qy 8161 GTCTCTCATACGATTTCAATACTCCCGACAGCGGTGAGTTCCTGCTGATTAOCT 8220
Db 8161 GTCTCTCATACGATTTCAATACTCCCGACAGCGGTGAGTTCCTGCTGATTAOCT 8220
Qy 8221 GGAATCAAGAAATGCCCTATGCGCTTCTCATATGACACCCGCTGTTGACTCAACGG 8280
Db 8221 GGAATCAAGAAATGCCCTATGCGCTTCTCATATGACACCCGCTGTTGACTCAACGG 8280
Qy 8281 TCACCTGAGAGTCACTTCGTGTGAGGAGTCAATTTACCAATGTTGACTGCGGCCCG 8340
Db 8281 TCACCTGAGAGTCACTTCGTGTGAGGAGTCAATTTACCAATGTTGACTGCGGCCCG 8340
Qy 8341 AGGCGAGACAGCGCTAAGTTCGTCTACAGAGCGCTTTACATCGGGGTCCCTGACTA 8400
Db 8341 AGGCGAGACAGCGCTAAGTTCGTCTACAGAGCGCTTTACATCGGGGTCCCTGACTA 8400
Qy 8401 ACTCAAAAGGGCAGAACTGCGGTATCGCGGTGCGCGAGTGCCTGCTGAGACTA 8460
Db 8401 ACTCAAAAGGGCAGAACTGCGGTATCGCGGTGCGCGAGTGCCTGCTGAGACTA 8460
Qy 8461 GCTGGGTAAATACCTTCATGTTTGAAGGCCACTGCAAGCTGCTGAGTCAAAAGC 8520
Db 8461 GCTGGGTAAATACCTTCATGTTTGAAGGCCACTGCAAGCTGCTGAGTCAAAAGC 8520
Qy 8521 TCCAGACTGCAAGTTCGTGTGAACGGAGAGCACTTGTGCTTATCTGTGAAGCGCGG 8580
Db 8521 TCCAGACTGCAAGTTCGTGTGAACGGAGAGCACTTGTGCTTATCTGTGAAGCGCGG 8580
Qy 8581 GAACCCAGGAGATGCGGGCGCTTACGAGCTTCAAGCGCTATGACTAGGTTATCCG 8640
Db 8581 GAACCCAGGAGATGCGGGCGCTTACGAGCTTCAAGCGCTATGACTAGGTTATCCG 8640
Qy 8641 CCCCCCGGGATCGGCCCAACAGAAATACGACTGAGCTGATAACATCATCTTCT 8700
Db 8641 CCCCCCGGGATCGGCCCAACAGAAATACGACTGAGCTGATAACATCATCTTCT 8700
Qy 8701 CCAATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8760
Db 8701 CCAATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8760
Qy 8761 CCACCAACCCCTTCCAGCGGCTGGGAGACAGTACACACTCCCACTCAACTCTT 8820
Db 8761 CCACCAACCCCTTCCAGCGGCTGGGAGACAGTACACACTCCCACTCAACTCTT 8820

QY 8821 GGCTAGGCAATATCATGATGATGCGCCACCCCTATGGGAAGGATGATTCGTGATGTC 8880
DB 8821 GGCTAGGCAATATCATGATGATGCGCCACCCCTATGGGAAGGATGATTCGTGATGTC 8880
QY 8881 ACTTTTCTCCATCCTTCTAGCTCAAGAGCAACTTGAAAAAGCCCTGGATGTCAGATCT 8940
DB 8881 ACTTTTCTCCATCCTTCTAGCTCAAGAGCAACTTGAAAAAGCCCTGGATGTCAGATCT 8940
QY 8941 ACGGGGCTGTACTCCATTGAGGCACTTGACCTACCTCAGATCAATGAAAGCTCCATG 9000
DB 8941 ACGGGGCTGTACTCCATTGAGGCACTTGACCTACCTCAGATCAATGAAAGCTCCATG 9000
QY 9001 GTCCTAGGCAATTTACACTCCACAGTACTCTCCAGGTGAGATCAATAGGGTGGCTCAT 9060
DB 9001 GTCCTAGGCAATTTACACTCCACAGTACTCTCCAGGTGAGATCAATAGGGTGGCTCAT 9060
QY 9061 GCCTAGGAAACTTGGGGTACCAACCTTGGAACTCGGAGACATCGGGCCAGAAAGTGTCC 9120
DB 9061 GCCTAGGAAACTTGGGGTACCAACCTTGGAACTCGGAGACATCGGGCCAGAAAGTGTCC 9120
QY 9121 GCGCTAGCTACTCTCCAGGGGGGAGGGGGGCACTTGTGGCAGATACCTCTTTAACT 9180
DB 9121 GCGCTAGCTACTCTCCAGGGGGGAGGGGGGCACTTGTGGCAGATACCTCTTTAACT 9180
QY 9181 GGGCAGTAAGGACCAAGCTTAAACTCACTCCAAATCCCGGGCCGCTCCAGCTGGACTGT 9240
DB 9181 GGGCAGTAAGGACCAAGCTTAAACTCACTCCAAATCCCGGGCCGCTCCAGCTGGACTGT 9240
QY 9241 CTGCTGGTTCGTGCTGTTACAGCGGGGAGACATATATCAAGCTGTCTCGTGCCC 9300
DB 9241 CTGCTGGTTCGTGCTGTTACAGCGGGGAGACATATATCAAGCTGTCTCGTGCCC 9300
QY 9301 GACCCCGTGTGTCGTTGTCCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9360
DB 9301 GACCCCGTGTGTCGTTGTCCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9360
QY 9361 TCCGCAACCGATGAACGGGGAGCTAAACCACTCCAGGCTTAAAGCCATTTCTGTTTTTT 9420
DB 9361 TCCGCAACCGATGAACGGGGAGCTAAACCACTCCAGGCTTAAAGCCATTTCTGTTTTTT 9420
QY 9421 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 9480
DB 9421 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 9480
QY 9481 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 9540
DB 9481 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 9540
QY 9541 CCGTGAGCGGATGACTGACAGAGTGTCTGATCTGCGCTCTCTGCGATCATGT 9595
DB 9541 CCGTGAGCGGATGACTGACAGAGTGTCTGATCTGCGCTCTCTGCGATCATGT 9595

RESULT 3
AAF23492
ID AAF23492 standard; DNA; 9595 Bp.
XX
AC AAF23492;
XX
DT 21-MAR-2001 (first entry)
XX
DE Infectious Hepatitis C virus 1b genotype.
XX
KW GBV-B; hepatitis C virus; HCV; vaccine; ds.
XX
OS Hepatitis C virus.
XX
PN WO200075337-A1.
XX
PD 14-DEC-2000.
XX
Ff 02-JUN-2000; 2000WO-US015293.

XX 04-JUN-1999; 99US-0137694P.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Bukh J, Yanagi M, Emerson SU, Purcell RH;
XX WPI; 2001-091214/10.
XX New infectious nucleic acids of the GB virus-B clone, useful for
PT indirectly studying the molecular properties of hepatitis C virus (HCV)
PT and in developing vaccines and therapeutics for HCV.
XX Disclosure; Fig 7; 96pp; English.
XX The present invention relates to GB virus-B. The nucleic acid molecules
CC of the invention are useful for indirectly studying the molecular
CC properties of hepatitis C virus (HCV). The infectious nucleic acid
CC sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used
CC in the development of vaccines and therapeutics for HCV
XX
SQ Sequence 9595 BP; 1934 A; 2843 C; 2697 G; 2121 T; 0 U; 0 Other;
Query Match 100.0%; Score 9593.4; DB 4; Length 9595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9594; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCAGCCCCCTGATGGGGGCGACACTCCACCATGAATCACTCCCTGTGAGGAAGTACTG 60
DB 1 GCCAGCCCCCTGATGGGGGCGACACTCCACCATGAATCACTCCCTGTGAGGAAGTACTG 60
QY 61 TCTTCACGAGAAAGCCTCTAGCCATGGGCTTAGTATGATGTCGTGTCAGCCTCCAGGAC 120
DB 61 TCTTCACGAGAAAGCCTCTAGCCATGGGCTTAGTATGATGTCGTGTCAGCCTCCAGGAC 120
QY 121 CCCCCCTCCCGGAGAGGCCATCTAGTGTCTCGGAACCGGTGAGTACACCGGAATTCGCAG 180
DB 121 CCCCCCTCCCGGAGAGGCCATCTAGTGTCTCGGAACCGGTGAGTACACCGGAATTCGCAG 180
QY 181 GACACCGGGTCTTCTTCTGGATCAACCCGCTCAATGCTCTGGAGATTTGGGCGTGCCCC 240
DB 181 GACACCGGGTCTTCTTCTGGATCAACCCGCTCAATGCTCTGGAGATTTGGGCGTGCCCC 240
QY 241 GCGAGACTGCTAGCCAGTAGTGTGGGTGCGAAGGCGCTTGTGTACTGCTGATAGG 300
DB 241 GCGAGACTGCTAGCCAGTAGTGTGGGTGCGAAGGCGCTTGTGTACTGCTGATAGG 300
QY 301 GTGCTTGCAGTCCCGCGGAGGTCTCGTAGACCGTGCACCATGAGCAGCAATCTTAAC 360
DB 301 GTGCTTGCAGTCCCGCGGAGGTCTCGTAGACCGTGCACCATGAGCAGCAATCTTAAC 360
QY 361 CTCAAGAAACCAACAGTAAACCAACCGCGCGCCACAGGACGTCAAGTTCCCGGGG 420
DB 361 CTCAAGAAACCAACAGTAAACCAACCGCGCGCCACAGGACGTCAAGTTCCCGGGG 420
QY 421 GTGCTCAGATCGTGTGGAGTATTACCTGTTCGCGGCGAGGCGCCAGGTTGGGTGTC 480
DB 421 GTGCTCAGATCGTGTGGAGTATTACCTGTTCGCGGCGAGGCGCCAGGTTGGGTGTC 480
QY 481 GCGCGACTAGGAAGGTTCGAGCGGTGCAACCTGTGGAAGGCGCAACCTATCCCAA 540
DB 481 GCGCGACTAGGAAGGTTCGAGCGGTGCAACCTGTGGAAGGCGCAACCTATCCCAA 540
QY 541 AGGCTCGCCGACCGGAGGCGGCGCTGGCTCAGCCCGGTTACCTTGGCCCTCTATG 600
DB 541 AGGCTCGCCGACCGGAGGCGGCGCTGGCTCAGCCCGGTTACCTTGGCCCTCTATG 600
QY 601 GCAATGAGGCGCTGGGGTGGGCGAGGATGGCTCTGTCAACCCCGCGCTCCCGGCTAGTT 660
DB 601 GCAATGAGGCGCTGGGGTGGGCGAGGATGGCTCTGTGTCAACCCCGCGCTCCCGGCTAGTT 660
QY 661 GGGGCCCCACGGAACCCCGGCTAGGTGGCTAACTTGGTAAAGGTATCATGATACCTTTA 720

2881 AATACTTTATCACAGAGCGGAGCCACATGCAAGTGTGGTCCGCCCTCAACGTTG 2940
Db
2881 AATACTTTATCACAGAGCGGAGCGCATGCAAGTGTGGTCCGCCCTCAACGTTG 2940
Qy
2941 GGGAGGCGCGATGCGATCATCTCTTCACTGTCGCGTTCAATCAGAGTTAAATTTTG 3000
Db
2941 GGGAGGCGCGCGATGCGATCATCTCTTCACTGTCGCGTTCAATCAGAGTTAAATTTTG 3000
Qy
3001 ACATCACCAAACTCTGTCGCGCATACCTCGGCGCCCTCATGTTGCTCCAGGCTGCATAA 3060
Db
3001 ACATCACCAAACTCTGTCGCGCATACCTCGGCGCCCTCATGTTGCTCCAGGCTGCATAA 3060
Qy
3061 CGAGAGTCGCTACTGTCGCGCTCAAGGCTCATTCGTCATGCAATGTTAGTGCAG 3120
Db
3061 CGAGAGTCGCTACTGTCGCGCTCAAGGCTCATTCGTCATGCAATGTTAGTGCAG 3120
Qy
3121 AAGTCGCGGGGCTCATTTATGTCGCAATGCTTTCATGAAGTGGGCGCTGACAGGTA 3180
Db
3121 AAGTCGCGGGGCTCATTTATGTCGCAATGCTTTCATGAAGTGGGCGCTGACAGGTA 3180
Qy
3181 CGTAGCTTTATAACCATCTTACCCCACTGCGGAGCTGGGCCACCGGCGCTACGAGACC 3240
Db
3181 CGTAGCTTTATAACCATCTTACCCCACTGCGGAGCTGGGCCACCGGCGCTACGAGACC 3240
Qy
3241 TTGCGGTGGCGTAGAGCCCGTCTTCTCCGCCATGGAGACCAAGGTCATCACCITGG 3300
Db
3241 TTGCGGTGGCGTAGAGCCCGTCTTCTCCGCCATGGAGACCAAGGTCATCACCITGG 3300
Qy
3301 GAGCAGACACGCTGCTGTCGCGCATCATCTTGGTCTACCGTCTCCGCCCGAAGG 3360
Db
3301 GAGCAGACACGCTGCTGTCGCGCATCATCTTGGTCTACCGTCTCCGCCCGAAGG 3360
Qy
3361 GGAAGGAGATATTTTGGACCGGCTGATGTCGAAAGGCAAGGTCGCGGCTCCCTG 3420
Db
3361 GGAAGGAGATATTTTGGACCGGCTGATGTCGAAAGGCAAGGTCGCGGCTCCCTG 3420
Qy
3421 CGCCCATCAGCGCTACTCCCAACAAACCGGGGCTACTTGGTGTGATCATCATAGCC 3480
Db
3421 CGCCCATCAGCGCTACTCCCAACAAACCGGGGCTACTTGGTGTGATCATCATAGCC 3480
Qy
3481 TCACAGGCGGGACAGAACAGGTCGAAGGGAGGTTCAAGTGGTTTCTACCGCAAC 3540
Db
3481 TCACAGGCGGGACAGAACAGGTCGAAGGGAGGTTCAAGTGGTTTCTACCGCAAC 3540
Qy
3541 AATCTTTCTGCGGACCTGCTATCAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
Db
3541 AATCTTTCTGCGGACCTGCTATCAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
Qy
3601 CGAAGACCTAGCGGTCCTCAAAAGTCCAAACCAAAATGACCAATGTAGACCTGG 3660
Db
3601 CGAAGACCTAGCGGTCCTCAAAAGTCCAAACCAAAATGACCAATGTAGACCTGG 3660
Qy
3661 ACCTGTCGCTGGCAGCGCGCCCGGGGCGCTCCATCAGACACCATGAGCTGTGCA 3720
Db
3661 ACCTGTCGCTGGCAGCGCGCCCGGGGCGCTCCATCAGACACCATGAGCTGTGCA 3720
Qy
3721 GCTCGGACCTTACTTGTGTGACGACATGCTGATGTCATTCGCGTGGCGGCGAGCG 3780
Db
3721 GCTCGGACCTTACTTGTGTGACGACATGCTGATGTCATTCGCGTGGCGGCGAGCG 3780
Qy
3781 ACAGAGGGAGTCTACTCTCCCGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3840
Db
3781 ACAGAGGGAGTCTACTCTCCCGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3840
Qy
3841 GTCCATGCTTTGCTTTCGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3900
Db
3841 GTCCATGCTTTGCTTTCGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3900
Qy
3901 GGGGGGTCGAGGCGGTGACCTTCAACCGCTTGAAGTCTATGCAAACTACCATGGGT 3960
Db
3901 GGGGGGTCGAGGCGGTGACCTTCAACCGCTTGAAGTCTATGCAAACTACCATGGGT 3960
Qy
3961 CTCGGTCTTTCACAGACACTCAACCCCGCGCTGTACCGAGACATTCGAAGTGGAC 4020

3961 CTCGGTCTTTCACAGACACTCAACCCCGCGCTGTACCGAGACATTCAGTGGAC 4020
Qy
4021 ATCTGACGCTCTTACTGTCAGCGGCAAGACACAAAGTCCGCTGCTGTCAGAGCC 4080
Db
4021 ATCTGACGCTCTTACTGTCAGCGGCAAGACACAAAGTCCGCTGCTGTCAGAGCC 4080
Qy
4081 AAGGTTACAAGTCTCTGTCGCAACCGTCTGTCGCGCACCTTGAAGTGGGGGT 4140
Db
4081 AAGGTTACAAGTCTCTGTCGCAACCGTCTGTCGCGCACCTTGAAGTGGGGGT 4140
Qy
4141 ATATGTCGAAGGCAACGTTATCCACCTCAATCAGAACTGGGTAAGGACCAATTAACA 4200
Db
4141 ATATGTCGAAGGCAACGTTATCCACCTCAATCAGAACTGGGTAAGGACCAATTAACA 4200
Qy
4201 CGGCGGCTCCATTAAGTCTTCCACCTATGCGCAAGTTCCTTGGCGACGCTGCTGCTG 4260
Db
4201 CGGCGGCTCCATTAAGTCTTCCACCTATGCGCAAGTTCCTTGGCGACGCTGCTGCTG 4260
Qy
4261 GGGGCGCTATGACATCATATATGTAAGTGGCACTCACTGAATCTGCACTACCATCT 4320
Db
4261 GGGGCGCTATGACATCATATATGTAAGTGGCACTCACTGAATCTGCACTACCATCT 4320
Qy
4321 TGGGATCGGCAACAGTCTGGAACAGCGGAGCGGTCGAGCGGCTGCTGCTGCTG 4380
Db
4321 TGGGATCGGCAACAGTCTGGAACAGCGGAGCGGTCGAGCGGCTGCTGCTGCTG 4380
Qy
4381 CCAACGCTTACCTCCGGATTCGCTTACCGTTCACACCCCAATATCGAGAAATAGCC 4440
Db
4381 CCAACGCTTACCTCCGGATTCGCTTACCGTTCACACCCCAATATCGAGAAATAGCC 4440
Qy
4441 TGTCCACATGAGAGATCCCTTCTATGGCAAGCAATCCCATTTGAGGCCATCAAG 4500
Db
4441 TGTCCACATGAGAGATCCCTTCTATGGCAAGCAATCCCATTTGAGGCCATCAAG 4500
Qy
4501 GGGGAGGCTCTCATTTTCTGCGATTCGCAAGAAATGAGAGCTCCGCCCAAGC 4560
Db
4501 GGGGAGGCTCTCATTTTCTGCGATTCGCAAGAAATGAGAGCTCCGCCCAAGC 4560
Qy
4561 TGAAGGCTCTGGAATGAAAGCTGATGATTAACCGGGGCTTGTGTCGCTGCTG 4620
Db
4561 TGAAGGCTCTGGAATGAAAGCTGATGATTAACCGGGGCTTGTGTCGCTGCTG 4620
Qy
4621 CGCTATCGGAGAGCTGCTGTCGTCGCAACAGAGCTCTAAAGAGGTTTCAACCGG 4680
Db
4621 CGCTATCGGAGAGCTGCTGTCGTCGCAACAGAGCTCTAAAGAGGTTTCAACCGG 4680
Qy
4681 ATTTTGAATCAGTGAATGCAATGATGTCGTCACCGGAGAGTGTGCTGCTGCTG 4740
Db
4681 ATTTTGAATCAGTGAATGCAATGATGTCGTCACCGGAGAGTGTGCTGCTGCTG 4740
Qy
4741 ATCCACCTTCAACATGAGAGAGAGCTGTCGTCGCAAGAGCGGCTGCTGCTGCTG 4800
Db
4741 ATCCACCTTCAACATGAGAGAGAGCTGTCGTCGCAAGAGCGGCTGCTGCTGCTG 4800
Qy
4801 GGGGAGTGAAGCTGGCAGGAGTGGAGTGGCATCTACAGGTTGTGCTGCTGCTGCTG 4860
Db
4801 GGGGAGTGAAGCTGGCAGGAGTGGAGTGGCATCTACAGGTTGTGCTGCTGCTGCTG 4860
Qy
4861 GGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
Db
4861 GGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
Qy
4921 GGTATGAGCTCAACCGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4980
Db
4921 GGTATGAGCTCAACCGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4980
Qy
4981 GGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5040
Db
4981 GGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5040
Qy
5041 ACATAGATGCCACTTCTGCTCCAGACTAAACAGGAGAGAGCACTTCTCTTACCTGG 5100

5041 ACATAGATGCCACTTCCTGTCCAGACTAAACAGGAGGAGCAACTTCCTTACTCGG 5100
5101 TGGCATATCAAGCTACAGTGTGGCCAGGCTCAAGCTCCACCTCCATCGTGGGAGCCAAA 5160
5101 TGGCATATCAAGCTACAGTGTGGCCAGGCTCAAGCTCCACCTCCATCGTGGGAGCCAAA 5160
5161 TGTGGAAGTGTCTCATACGGCTGAACCTACACTGACAGGCGCAACACCCCTGCTGTATA 5220
5161 TGTGGAAGTGTCTCATACGGCTGAACCTACACTGACAGGCGCAACACCCCTGCTGTATA 5220
5221 GGCTAGGAGCGCTCCAAAATGAGGTGATCTCTACACACCCCAATAACTAAATACATATGG 5280
5221 GGCTAGGAGCGCTCCAAAATGAGGTGATCTCTACACACCCCAATAACTAAATACATATGG 5280
5281 CATGCATGTCCGGCTGACCTCGAGGTGCTCTACAGCACTGGGTGCTGGTAGGGCGGAGTCC 5340
5281 CATGCATGTCCGGCTGACCTCGAGGTGCTCTACAGCACTGGGTGCTGGTAGGGCGGAGTCC 5340
5341 TTGCAGCTTTGGCGGCACTACTGCTGACGACAGGCAAGTGTGGTCAATTTGGGCGAGATCA 5400
5341 TTGCAGCTTTGGCGGCACTACTGCTGACGACAGGCAAGTGTGGTCAATTTGGGCGAGATCA 5400
5401 TCTTTCCGGGAAGCCAGCTGTCTTCCGACAGGGAAGTCTCTACACAGGAGTTCGATG 5460
5401 TCTTTCCGGGAAGCCAGCTGTCTTCCGACAGGGAAGTCTCTACACAGGAGTTCGATG 5460
5461 AGATGAAGAGTGTGCTCTCAAACTTCTTACATCGAGCAGGGAATGCAGCTGCCCGAGC 5520
5461 AGATGAAGAGTGTGCTCTCAAACTTCTTACATCGAGCAGGGAATGCAGCTGCCCGAGC 5520
5521 AATTCAAGCAAAAGGCGCTCGGTTGTGCAAAAGGCGCACCAAGAGCGGAGGCTGCTG 5580
5521 AATTCAAGCAAAAGGCGCTCGGTTGTGCAAAAGGCGCACCAAGAGCGGAGGCTGCTG 5580
5581 CTCCTGTGTGAGTCCAAAGTGGCGAGCCCTTCAGACCTTCTGGGCGAACAATGGA 5640
5581 CTCCTGTGTGAGTCCAAAGTGGCGAGCCCTTCAGACCTTCTGGGCGAACAATGGA 5640
5641 ATTTCTACAGCGAATACAGTACTAGCAGGCTTATCCACTCTGCTTGGAAACCCCGCGA 5700
5641 ATTTCTACAGCGAATACAGTACTAGCAGGCTTATCCACTCTGCTTGGAAACCCCGCGA 5700
5701 TAGCATCATGATGCGATTTACAGCTTCTATCACTAGCCGCTCACCAACCAAAAACCC 5760
5701 TAGCATCATGATGCGATTTACAGCTTCTATCACTAGCCGCTCACCAACCAAAAACCC 5760
5761 TCTGTGTTAACTTTGGGGGATGGTGGCTGCCAACTCGCTCTCCAGCGCTGCGT 5820
5761 TCTGTGTTAACTTTGGGGGATGGTGGCTGCCAACTCGCTCTCCAGCGCTGCGT 5820
5821 CAGCTTCGTGGGCGCGGCATCGCGAGCGGCTGTGGCAGCATAGGCTTTGGGAAG 5880
5821 CAGCTTCGTGGGCGCGGCATCGCGAGCGGCTGTGGCAGCATAGGCTTTGGGAAG 5880
5881 TGTCTGTGACATCTTTGGGGGATGGTGGCTGCCAACTCGCTCTCCAGCGCTGCGT 5940
5881 TGTCTGTGACATCTTTGGGGGATGGTGGCTGCCAACTCGCTCTCCAGCGCTGCGT 5940
5941 AGGTCTATGAGCGGAGGTGCCCTCCACCGAGGACCTGTCAACTTACTCCCTGCCATCC 6000
5941 AGGTCTATGAGCGGAGGTGCCCTCCACCGAGGACCTGTCAACTTACTCCCTGCCATCC 6000
6001 TCTCTCTGTGACCTGTGCTGGGTCGTGTGCGGAGCAATPACTGCGTTCGCGAGTGA 6060
6001 TCTCTCTGTGACCTGTGCTGGGTCGTGTGCGGAGCAATPACTGCGTTCGCGAGTGA 6060
6061 GCCCGGAGAGGGGCTGTGCTGAGTGAACCGGCTGATAGGCTTCGCTTCGCGGGTA 6120
6061 GCCCGGAGAGGGGCTGTGCTGAGTGAACCGGCTGATAGGCTTCGCTTCGCGGGTA 6120
6121 ACCACGTCTCCCTACGCACTATGCTGCTGAGAGCGAGCGCTGAGGACGCTGCTACATCAGA 6180
6121 ACCACGTCTCCCTACGCACTATGCTGCTGAGAGCGAGCGCTGAGGACGCTGCTACATCAGA 6180

6181 TCCCTCTAGCCCTTACCATCACTCAACTGCTGAAGGGCTCCACCAGTGGATTAATGAGG 6240
6181 TCCCTCTAGCCCTTACCATCACTCAACTGCTGAAGGGCTCCACCAGTGGATTAATGAGG 6240
6241 ACTGTCTACGCCATGCTCCGGCTCGTGGCTAAGGATGTTTGGATGAGATGACACGG 6300
6241 ACTGTCTACGCCATGCTCCGGCTCGTGGCTAAGGATGTTTGGATGAGATGACACGG 6300
5301 TGTGACTGACCTTCAGACCTGGCTCCAGTCCAACTCCTGCGCGGTTTACCGGAGTCC 6360
5301 TGTGACTGACCTTCAGACCTGGCTCCAGTCCAACTCCTGCGCGGTTTACCGGAGTCC 6360
6361 CTTTCTGTCTATGCCAAACGCGGGTACAGGGAGTGTGGCGGGGAGCGCATCATGCAAA 6420
6361 CTTTCTGTCTATGCCAAACGCGGGTACAGGGAGTGTGGCGGGGAGCGCATCATGCAAA 6420
6421 CCACTGTGCCATCGCGAGCACAGATCGCGGACATGTCAAAAACGGTTCCTCATGAGATCG 6480
6421 CCACTGTGCCATCGCGAGCACAGATCGCGGACATGTCAAAAACGGTTCCTCATGAGATCG 6480
6481 TAGGGCTAGAACCTGCAGACACAGTGGCACGGACGTTCCCATCAACGATACACA 6540
6481 TAGGGCTAGAACCTGCAGACACAGTGGCACGGACGTTCCCATCAACGATACACA 6540
5541 CGGGACCTTGCACACCTCCCGCGGCCCACTATTCCAGGGCGCTATGGCGGGTGGTG 6600
5541 CGGGACCTTGCACACCTCCCGCGGCCCACTATTCCAGGGCGCTATGGCGGGTGGTG 6600
5601 CTGAGAGTACGTGAGGTTACGCTGTGGGGATTTCCACTACGTGACGGCATGACCA 6660
5601 CTGAGAGTACGTGAGGTTACGCTGTGGGGATTTCCACTACGTGACGGCATGACCA 6660
6661 CTGACAAAGTAAAGTCCCATGCGAGTTCGGGCCCGGAACTTCTTACGGAGGTGGATG 6720
6661 CTGACAAAGTAAAGTCCCATGCGAGTTCGGGCCCGGAACTTCTTACGGAGGTGGATG 6720
6721 GAGTGGGTGACACAGGTACGCTCCGGGTGCAAACTCTTCTACGGAGGACGTCAGT 6780
6721 GAGTGGGTGACACAGGTACGCTCCGGGTGCAAACTCTTCTACGGAGGACGTCAGT 6780
6781 TCCAGTGGGGTCAACCAATCTTGTGGGTGCGAGTCCCATGCGCGCCGAAACCGG 6840
6781 TCCAGTGGGGTCAACCAATCTTGTGGGTGCGAGTCCCATGCGCGCCGAAACCGG 6840
6841 ACGTAAAGTGTCTTCTTCCATGCTCACGATCCCTCCACATACAGAGAGCGGCTA 6900
6841 ACGTAAAGTGTCTTCTTCCATGCTCACGATCCCTCCACATACAGAGAGCGGCTA 6900
6901 AGCGTAGGCTGAGAGGTCTCCCTCTCTTTAGCCAGGTCTCATCAGTACGCACTTGT 6960
6901 AGCGTAGGCTGAGAGGTCTCCCTCTCTTTAGCCAGGTCTCATCAGTACGCACTTGT 6960
6961 CTGCGCTCTTTTGAAGGAGCATGCACTACCCACATGCTCCCGGAGCGTCACTCA 7020
6961 CTGCGCTCTTTTGAAGGAGCATGCACTACCCACATGCTCCCGGAGCGTCACTCA 7020
7021 TCCAGGCCAACTCTTGTGGGCGAGGATGGGCGGAAACATCTCCTGCTGAGTCA 7080
7021 TCCAGGCCAACTCTTGTGGGCGAGGATGGGCGGAAACATCTCCTGCTGAGTCA 7080
7081 AGAATAAGTGTAAATCTTGGACTCTTTTGAACCGCTTTCAGCGAGGGGATGAGAGG 7140
7081 AGAATAAGTGTAAATCTTGGACTCTTTTGAACCGCTTTCAGCGAGGGGATGAGAGG 7140
7141 AGATATCGTTCGCGCGGAGATCTTCCACTGCTAGAGTCCCTGAGGAGCGGACTACG 7200
7141 AGATATCGTTCGCGCGGAGATCTTCCACTGCTAGAGTCCCTGAGGAGCGGACTACG 7200
7201 TATGGGACGCGCGGACTTACATCTCCACTGCTAGAGTCTTGGNAGGACCGGACTACG 7260
7201 TATGGGACGCGCGGACTTACATCTCCACTGCTAGAGTCTTGGNAGGACCGGACTACG 7260

QY	7261	TCCCTCCGGTGGTACACGGATGCCCATTTGCCACCTTACCAAGGCTCTCCCAATACCACTC	7320
DB	7261	TCCCTCCGGTGGTACACGGATGCCCATTTGCCACCTTACCAAGGCTCTCCCAATACCACTC	7320
QY	7321	CACGGAGAAAGAGACGGTGTCTCTGACAGAAATCCAAATGTGTCTTCTGCCTTGGCGGAGC	7380
DB	7321	CACGGAGAAAGAGACGGTGTCTCTGACAGAAATCCAAATGTGTCTTCTGCCTTGGCGGAGC	7380
QY	7381	TGCGCACTAAGACCTTCGGTAGCTCCGGATCCTCGGCCGTGTGATAGCGGCACGGCCACCG	7440
DB	7381	TGCGCACTAAGACCTTCGGTAGCTCCGGATCCTCGGCCGTGTGATAGCGGCACGGCCACCG	7440
QY	7441	CCCTTCCTGACCTGGCCTCCGACGACGGTGAACAAAGGATCCGAGCTTGAGTCTCTCT	7500
DB	7441	CCCTTCCTGACCTGGCCTCCGACGACGGTGAACAAAGGATCCGAGCTTGAGTCTCTCTCT	7500
QY	7501	CCATGCCCCCTTTGAAGGGGAGCCGGGGACCCCGATCTCAGCGACGGGTCTTGCTCTA	7560
DB	7501	CCATGCCCCCTTTGAAGGGGAGCCGGGGACCCCGATCTCAGCGACGGGTCTTGCTCTA	7560
QY	7561	CCGTGAGTGAGAGGCTAETGAGGATGTCTGTCTGCTCTCAATGTCTTATACGTGGACAG	7620
DB	7561	CCGTGAGTGAGAGGCTAETGAGGATGTCTGTCTGCTCTCAATGTCTTATACGTGGACAG	7620
QY	7621	GCGCCCTGATCACGCCATGCGCTCGGAGGAAATGAAGCTGCCCATCAACCCGTTGAGCA	7680
DB	7621	GCGCCCTGATCACGCCATGCGCTCGGAGGAAATGAAGCTGCCCATCAACCCGTTGAGCA	7680
QY	7681	ACTCTTTGCTCGCTCACCAAAATGCTACGCCCAACATCCGCGAGCGCAAGCCTCC	7740
DB	7681	ACTCTTTGCTCGCTCACCAAAATGCTACGCCCAACATCCGCGAGCGCAAGCCTCC	7740
QY	7741	GGCAGAGAAGGTCACTTTACAGATTTGCAAGTCTCTGGATGATCATTCACCGGAGCTAC	7800
DB	7741	GGCAGAGAAGGTCACTTTACAGATTTGCAAGTCTCTGGATGATCATTCACCGGAGCTAC	7800
QY	7801	TCAAGGAGATGAAGCGAAGCGGTCACAGTTAAGCTAAAGCTTCTATCTATAGAGGAGG	7860
DB	7801	TCAAGGAGATGAAGCGAAGCGGTCACAGTTAAGCTAAAGCTTCTATCTATAGAGGAGG	7860
QY	7861	CCTGCAAGCTGACGCCCCACATTCGGCCAAAATCCAAAATTTGGCTATGGGGCAAGGACG	7920
DB	7861	CCTGCAAGCTGACGCCCCACATTCGGCCAAAATCCAAAATTTGGCTATGGGGCAAGGACG	7920
QY	7921	TCCGGAACTTATCCAGCAGGCGCGTTAAACACATCCGCTCCGTGTGGAGGACTTCTCTG	7980
DB	7921	TCCGGAACTTATCCAGCAGGCGCGTTAAACACATCCGCTCCGTGTGGAGGACTTCTCTG	7980
QY	7981	AAGACACTGAACACACAAATTGACACCAATCATATGCGAAAAGTAGAGTTTCTCGCTCC	8040
DB	7981	AAGACACTGAACACACAAATTGACACCAATCATATGCGAAAAGTAGAGTTTCTCGCTCC	8040
QY	8041	AACACAGAAAGGAGGCGCGCAAGCCAGCTCGCTTATCGTATTTCCAGACCTGGGAGTTC	8100
DB	8041	AACACAGAAAGGAGGCGCGCGCAAGCCAGCTCGCTTATCGTATTTCCAGACCTGGGAGTTC	8100
QY	8101	GTGTATGCGAAGAAGTGGCCCTTTACGACGTGTCTCCACCTTCTCAGGCGGTATGCG	8160
DB	8101	GTGTATGCGAAGAAGTGGCCCTTTACGACGTGTCTCCACCTTCTCAGGCGGTATGCG	8160
QY	8161	GCTCTCTATACGATTTCAATATCCGCCAAGCAGCGGGTTCGAGTTTCTGGTGAATACCT	8220
DB	8161	GCTCTCTATACGATTTCAATATCCGCCAAGCAGCGGGTTCGAGTTTCTGGTGAATACCT	8220
QY	8221	GGAAATCAAGAAGTGCCTATGGCTTCTCATATGACACCCGCTCTTTTGTACTCAACGG	8280
DB	8221	GGAAATCAAGAAGTGCCTATGGCTTCTCATATGACACCCGCTCTTTTGTACTCAACGG	8280
QY	8281	TCACTGAGAGTGACATTCGTGTTTGAGGAGTCAATTTTACCAAATGTTGTGACTTGGCCCCG	8340
DB	8281	TCACTGAGAGTGACATTCGTGTTTGAGGAGTCAATTTTACCAAATGTTGTGACTTGGCCCCG	8340
QY	8341	AGGCCAGACAGGCCATAAGGTCTGCTCACAGAGCGGCTTTACATCGGGGGTCCCTGACTA	8400

[illegible]

Db 1022 GGAGAGTAATTTCTCCCGTTGCTGGGTAGCGCTCACTCCACGCTCGCGCCAGGAACAG 1081
Qy 1094 CAGCGTCCCGCACTACGAAATACGACGCGCACGTCGACTTTCGCTGCTGGGACGCGTCTTT 1153
Db 1082 CAGCATCCCGCACACGACAAATAGGACGACAGCTGCTGCTGCTGGGCGGCTGCTCT 1141
Qy 1154 CTGCTCCGCTAATGTAAGTGGGGGATCTCTGCGGAATCTATTTTCCTCGTCCCGACTGTT 1213
Db 1142 CTGTTCCGCTAATGTAAGTGGGGGATCTCTGCGGAATCTCTGCTCGTCCCGACTGTT 1201
Qy 1214 CACCTTCTCGCTCCGCGCATGAGACAGTGCAGAGCTGCACACTGCTCACTATTCOCGG 1273
Db 1202 CACCTTCTCACTCCGCGGTATGAGACGATCAAGATTGCAATTGCTCAATCTATCCCGG 1261
Qy 1274 CCAATGATCAGTCCACCGCATGGCTTGGGATGATGATGAATGAACTGGTCACTTACACACG 1333
Db 1262 CCACTGATCAGTCCACCGCATGGCTTGGGATGATGATGAATGAACTGGTCACTTACACGCG 1321
Qy 1334 CTAAGTGTGTCGAGTTGCTCCCGATCCCAAGCTGTCGTTGGACATGGTGGCGGGGCG 1393
Db 1322 CTAAGTGTGTCGAGTTGCTCCCGATCCCAAGCTGTCGTTGGACATGGTGGCGGGGCG 1381
Qy 1394 CCACGTGGGAGTCCGCGGGCTTGGCTTACTATTCATGCTAGGAACTGGCTAAAGT 1453
Db 1382 CCACGTGGGAGTCCGCGGGCTTGGCTTACTATTCATGCTAGGAACTGGCTAAAGT 1441
Qy 1454 TCTGATTTGGCGCTACTCTTTGCGCGGTTGACGCGGAGACCCACACGACGCGGAGGT 1513
Db 1442 CTTGATTTGATGCTTACTTGTGCTGGCTTACGCGGCACACCGCTGACAGGGGAG 1501
Qy 1514 GCGCGGCACACACTCCGGTTCACTGCTCTTTTCACTGGGGGCTCAGAAAT 1573
Db 1502 GGTAGCTCCAGCACCCAGAGCTCGTGTCTGGCTCTCACAAAGCCCATCTCAGAAAT 1561
Qy 1574 CCAGCTTGTGAATCCAAAGCGAGCTGGACATCAACAGGCTGCGCTAAATTCGAATGA 1633
Db 1562 CCACGTGTGAACCAACAGCGAGTGGACATCAACAGGACCGCTCTGAATTCGAATGA 1621
Qy 1634 CTCCTCCAAACTGGGTTCTTTGCGCGCTGTTTACGACACAAAGTTCAACTGCTCGG 1693
Db 1622 CTCCTCCAAACTGGGTTCTTTGCGCGCTGTTTACGACACAAAGTTCAACTGCTCGG 1681
Qy 1694 GTGCGCGAGCGATGGCGAGCTGCGGCCCATTTGACTGGTTGCGCCAGGGGTGGGGCC 1753
Db 1682 GTGCGCGAGCGATGGGATGCTGCGGCCCATTCGATGAGTTGCTCAGGGGTGGGGTCC 1741
Qy 1754 CATCACCTTATTAAGCTTAAGCTGCGATCAGAGGCTTATTTGCTGCGATTACGCGCC 1813
Db 1742 CATCACTCATGATGCTTGAGCTCGGACAGAGGCTTATTTGCTGCGACTACGCGCC 1801
Qy 1814 TCGACCGTGTGTPCGTACCGCGCTGCGAGTGTGCTGCTCAGTGTATTTGTTTCAACCCC 1873
Db 1802 TCGACCGTGTGTPCGTACCGCGCTGCGAGTGTGCTGCTCAGTGTATTTGCTTCACTCC 1861
Qy 1874 AAGCCCTGTTGTTGGGGACACCGATCGTTCGGTGTCCCTAGCTATAGTGGGGGA 1933
Db 1862 GAGCCCTGTTGTTGGGGACACCGATCGTTCGGGCTCCTACGATAGTGGGGGA 1921
Qy 1934 GAAATGAGACAGACGTGATGCTCTCAACACACGCGTCCGCCACAAAGGCAACTGTTTCGG 1993
Db 1922 GAAATGAGACAGACGTGATGCTCTCAACACACGCGGCTCCGCTCAGGCAACTGTTTCGG 1981
Qy 1994 CTGTACATGGATGATAGTACTGGGTTCACTAAGACGTGCGAGTCCCGCTGTAAAT 2053
Db 1982 CTGTACATGGATGATAGTACTGGGTTCACTAAGACGTGCGAGGCTCCCGTGTAAAT 2041
Qy 2054 CGGGGGGTGCGTACCGCACTTCACTGCTCCCGAGCTGCTTCCGGAAGCAACCCCGA 2113
Db 2042 CGGGGGGTGCGGCAACCACTTGGTGTGCGCCCAAGGATGCTTCCGGAAGCAACCCCGA 2101
Qy 2114 GGCCTATTACAAATCTGGCTCGGGGCTTGGTTGACACTAGTGGTCTAGTACTA 2173
Db 2102 GGCCTATTACAAATCTGGCTCGGGGCTTGGTTGACACTAGTGGTCTAGTACTA 2161

Qy 2174 CCCATACAGGCTTTGGGCACTACCCCTGCACTCTCAATTTTCCATCTTTAGGTTAGAT 2233
Db 2162 CCCATACAGGCTTTGGGCACTACCCCTGCACTCTCAATTTTCCATCTTTAGGTTAGAT 2221
Qy 2234 GTATGTGGGGGCTGGAGCACAGGCTCAATSCCGCATGCAATTOGACTCGAGGAGAGCG 2293
Db 2222 GTATGTGGGGGCTGGAGCACAGGCTCAATGCTGATGCAATTTGGAATCGAGGAGAGCG 2281
Qy 2294 CTGTAACTTTGAGAGACAGGATAGTTCAGAACTCAGCCCGCTGCTGTCTCAACAGA 2353
Db 2282 CTGTACTTTGAGAGACAGGATAGTTCAGAACTCAGCCCGCTGCTGTCTCAACAGA 2341
Qy 2354 GTGGCAGATACTGCGCTGTGCTTTCACCAACCTTACCGGCTTTATCCACTGCTTTGATCCA 2413
Db 2342 GTGGCAGATACTGCGCTGTGCTTTCACCAACCTTACCGGCTTTATCCACTGCTTTGATCCA 2401
Qy 2414 TCTCCATCAGAACTCGTGGACGTGCAATACCTGTACGTTAGGTCAGGCTTTGCTC 2473
Db 2402 TCTTCAACGGAACATCGTGGACGTGCAATACCTGTACGTTAGGTCAGGCTTTGCTC 2461
Qy 2474 CTTTGCATCAAAATGGGAGTACATCTGTTGCTTTTCTCTCTCTGGAGACGCGCGGT 2533
Db 2462 CTTTGCATCAAAATGGGAGTACATCTGTTGCTTTTCTCTCTCTCTCTCTCTCTCTCT 2521
Qy 2534 GTGTGCTCTGCTTGTGGATGATGCTGCTGATAGCCAGGCTGAGCGCGCTTTAGAGAACT 2593
Db 2522 CTGTGCTCTGCTTGTGGATGATGCTGCTGATAGCCAGGCTGAGCGCACCTTTAGAGAACT 2581
Qy 2594 GGTGCTCTCAATCGCGGCTCGTGGCGGAGCATGATTTCTCTCTCTCTCTCTCTCTCT 2653
Db 2582 GGTGCTCTCAATCGCGGCTCGTGGCGGAGCATGATTTCTCTCTCTCTCTCTCTCTCT 2641
Qy 2654 CTTCTGCGCGCTCTGTTACATTAAGGCGAGGCTGCTGCTCTCTCTCTCTCTCTCTCTCT 2713
Db 2642 CTTCTGCGCGCTCTGTTACATTAAGGCGAGGCTGCTGCTCTCTCTCTCTCTCTCTCTCT 2701
Qy 2714 TGGGCTATGGCGCTGCT 2773
Db 2702 TGGGCTATGGCGCTGCT 2761
Qy 2774 CCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2833
Db 2762 CCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2821
Qy 2834 GTCAACCATCTACAACTGTTTCTCACTAGGCTCATATGCTGCTGCTGCTGCTGCTGCT 2893
Db 2822 GTCAACCATCTACAACTGTTTCTCACTAGGCTCATATGCTGCTGCTGCTGCTGCTGCT 2881
Qy 2894 CAGAGCGAGCGCACATGCAAGTGTGGGTCTCCCGCTCAACGCTTCCGGGAGCGCGCGA 2953
Db 2882 CAGAGCGAGCGCACATGCAAGTGTGGGTCTCCCGCTCAACGCTTCCGGGAGCGCGCGA 2941
Qy 2954 TGCATCATCT 3013
Db 2942 TGCATCATCT 3001
Qy 3014 CTTCTCTGCGCATCTCTGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3073
Db 3002 CTTCTCTGCGCATCTCTGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3061
Qy 3074 CTTCTCTGCGCGCTCAAGGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3133
Db 3062 CTTCTCTGCGCGCTCAAGGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3121
Qy 3134 TCATTAATGCTCAAAATGCTTTCATGAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 3193
Db 3122 CCACTATGCTCAAAATGCTTTCATGAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3181
Qy 3194 CCACTATGCTCAAAATGCTTTCATGAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3253
Db 3182 CCACTATGCTCAAAATGCTTTCATGAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3241

QY 3254 AGAGCCCGTCTCTCCGCCCATGAGACCAAGGTCATCACCTGGGGGAGACACCCG 3313
 DB 3242 AGAGCCCGTCTCTCTGACATGAGACCAAGGTCATCACCTGGGGGAGACACCCG 3301
 QY 3314 TCGTGTGGGGAATCATCTTGGGTCTACCCGTCTCCGCCGAGGGGAGAGATATTT 3373
 DB 3302 GGGGTGTGGGGAATCATCTCGGTCTACCCGTCTCCGCCGAGGGGAGAGATATTT 3361
 QY 3374 TTTGGGACCGGTGTATGTCCTCAAGGGCAAGGGTGGGACTCTCTGGGCCATCCGGC 3433
 DB 3362 TCTAGGACCGGCGATATTTTGGAGAGAGGGTGGGCTCTCTGGGCCATCCGGC 3421
 QY 3434 CTACTCCCAACAAACGGGGGCGTACTTGGTTCATCATCTAGGCTCAACGCGGGA 3493
 DB 3422 CTATTCCCAACAAACGGGGGCGTCTGGTTCATCATCTAGGCTCAACGCTCGGA 3481
 QY 3494 CAGAACCCAGTTCGAGGGGAGGTTCAATGGTTCATCCGCAACCAATCTTCTGGC 3553
 DB 3482 CAAGAACCCAGTTCGAGGGGAGGTTCAATGGTTCATCCGCAACCAATCTTCTGGC 3541
 QY 3554 GACTTGATCAACGGGTGTCTGAGCTGTCTACCATGGGCTCGAGAGACCTTAGC 3613
 DB 3542 GACTGGTCAATGGGCTGTCTGAGCCGTCTACCATGGGCTCGAGAGACCTTAGC 3601
 QY 3614 CGTCCAAAGGTCGAATCAACCAATGTACCAATGTAGACCTGGACCTCGTGGCTG 3673
 DB 3602 CGGCGCAAGGGTCCAATCAACCAATGTACCAATGTAGACCACTCGTGGCTG 3661
 QY 3674 GCGGGGCCCCCGGGGCGGCTCCATGACCAATGAGAGTGTGGAGCTCGGACCTTGA 3733
 DB 3662 GCGGGGCCCCCGGGGCGGCTCCATGACCACTGACCGCTCGGAGCTCGGACCTTGA 3721
 QY 3734 CTGTGTCACGAGCATGCTGATGTCATTCGGGTGCGCGGCGAGGCGACAGCAGGGGAG 3793
 DB 3722 CTGTGTCACGAGCATGCTGATGTCATTCGGGTGCGCGGCGAGGCGACAGCAGGGGAG 3781
 QY 3794 TCTACTCTCCCGAGGCGGCTCTCTCACTGAAAGGCTCTCGGGTGGTCCAATGCTTTG 3853
 DB 3782 CTGTCTTCCCGAGGCGGCTCTCTCACTGAAAGGCTCTCGGGTGGTCCAATGCTTTG 3841
 QY 3854 CCTTCGGGACGCTGTGGGCGTCTTCCGGGCTGTGTGTGACCGGGGGGTGCGGA 3913
 DB 3842 CCTTCGGGACGCTGTGAGGCACTTCCGGGCTGTGTGTGACCGGGGGGTGCGGA 3901
 QY 3914 GCGGTGGACTCTCATACCCGCTGAGTCTATGGAATACCATCGGTCTCCGGTCTTAC 3973
 DB 3902 GCGGTGGACTCTCATACCCGCTGAGTCTATGGAATACCATCGGTCTCCGGTCTTAC 3961
 QY 3974 AGACACTCAACCCCGGCTGTACGCGAGACATTCCAAGTGGCAATCTGCAAGCTCC 4033
 DB 3962 AGACACTCAATCCCGCTTCCGGGCTGTGTGTGACCGGGGGGTGCGGA 4021
 QY 4034 TACTGGCAGCGGCAAGAGCACAAGTGGCGGCTGTGTGTGACCGGGGGGTGCGGA 4093
 DB 4022 CATGGCAGCGGCAAGAGCACAAGTGGCGGCTGTGTGTGACCGGGGGGTGCGGA 4081
 QY 4094 GCTCGTCTGAACCCGCTCGGTTCGCGCACTTAGGGTTTGGGGCGGTATATGTCGAAGC 4153
 DB 4082 GCTCGTCTGAACCCGCTCGGTTCGCGCACTTAGGGTTTGGGGCGGTATATGTCGAAGC 4141
 QY 4154 ACAGGTATCGACCTCAATCACTGAGTGGGTGAGGACATTCACCGGGGCGGTCCAT 4213
 DB 4142 ACATGGCATCGACCTCAATCACTGAGTGGGTGAGGACATTCACCGGGGCGGTCCAT 4201
 QY 4214 TAGCTACTCCACCTATGCGAGTCTCTGGGAGGTTGGGTCTCTGGGGGCGGTATGA 4273
 DB 4202 CAGCTACTCCACCTATGCGAGTCTCTGGGAGGTTGGGTCTCTGGGGGCGGTATGA 4261
 QY 4274 CATCATATATGATGAGTGGCACTCACTGAGTCCAGTACCATCTTGGGCTCGGCAC 4333
 DB 4262 CATCATATATGATGAGTGGCACTCACTGAGTCCAGTACCATCTTGGGCTCGGCAC 4321
 QY 4334 AGTCTCGACCAAGCGGAGAGCGGTGGAGCGGGCTCGTCTGCCACCGCTACCC 4393

DB 4322 AGTCTCGATCAGGCAGAGACGGCTGGAGCGGCTCGTGTGTCGCCACCGCCACGCC 4381
 QY 4394 TCGGGATCGGTACCGTGGCGACACCCCAATATCGAGGAATAGSCCTGTCCAAAGG 4453
 DB 4382 TCGGGATCGATCACCGTGGCGACACCCCAATATCGAGGAATAGSCCTGTCCAAAGG 4441
 QY 4454 AGAGATCCCTTCTATGCGCAAGCCATCCCATTCAGGTCATCAAGGGGGGAGGCACT 4513
 DB 4442 AGAGATCCCTTCTATGCGCAAGCCATCCCATTCAGGTCATCAAGGGGGGAGGCACT 4501
 QY 4514 CATTTCTGCOATTCOAAGAAATGAGAGCTCGCGCAAGCTGACAGGCTCG 4573
 DB 4502 CATCTTCCCATTCOAAGAAATGAGAGCTCGCGCAAGCTGACAGGCTCG 4561
 QY 4574 ACTGACCTGTAGCATATTACGGGGCTTGTATGTCGGTCAACCGCTATCGGAGA 4633
 DB 4562 ACTCAATGCTGAGCGTATTACGGGGCTTGTATGTCGGTCAACCGCTATCGGAGA 4621
 QY 4634 CGTCTGTGCGGGCAACAGACCTCTAATGACGGGTTTCAACGGCGATTTTGACTCAGT 4693
 DB 4622 CGTCTGTGCGGGCAACAGACCTCTAATGACGGGTTTACCGGCGACTTTGACTCAGT 4681
 QY 4694 GATGACTGCAATACATGTGTCAACAGAGCTGCACTTCACTTGGATTCACCTTAC 4753
 DB 4682 GATGACTGCAATACATGTGTCAACAGAGCTGCACTTCACTTGGATTCACCTTAC 4741
 QY 4754 CATTCAGAGACAGCGTCCCGAGAGCGGTGTGCGCTCGCAACCGCGAGGTAGAAC 4813
 DB 4742 CATTCAGAGACAGCGTCCCGAGAGCGGTGTGCGCTCGCAACCGCGAGGTAGAAC 4801
 QY 4814 TGGCGAGGCTAGGAGTGGCATCTACAGGTTTGTGACTCCAGAGAACCGGCTCGGGAT 4873
 DB 4802 TGGCGAGGCTAGGAGTGGCATCTACAGGTTTGTGACTCCAGAGAACCGGCTCGGGAT 4861
 QY 4874 GTTCGATTTCTCGGTCTGTGTGAGTGTGCTATGACGGGGCTGTGTGTGATGACTCAC 4933
 DB 4862 GTTCGATTTCTCGGTCTGTGTGAGTGTGCTATGACGGGGCTGTGTGTGATGACTCAC 4921
 QY 4934 GCGCGTGAACCTCGGTAGGTTCGGGCTTACCTTAATACACAGGGTTCCCGGCTG 4993
 DB 4922 GCGCGTGAACCTCGGTAGGTTCGGGCTTACCTTAATACACAGGGTTCCCGGCTG 4981
 QY 4994 CCAGGACCATCTGGAGTCTGGGAGAGCTCTTCACAGGCTTCACCCACATAGATGCCA 5053
 DB 4982 CCAGGACCATCTAGAGTCTGGGAGAGCTCTTCACAGGCTTCACCCACATAGATGCCA 5041
 QY 5054 CTTCGTGTCGAGCTAATAACAGGCGAGAGAACCTTCTTACCTTGGTGGCATATCAAGC 5113
 DB 5042 CTTCGTGTCGAGCTAATAACAGGCGAGAGAACCTTCTTACCTTGGTGGCATATCAAGC 5101
 QY 5114 TACAGTGTGCGCGAGGCTCAAGCTCCACCTCCATGTTGGGACCAATGTGGAGTGTCT 5173
 DB 5102 CACAGTGTGCGCGAGGCTCAGCTCCACCTCCATGTTGGGACCAATGTGGAGTGTCT 5161
 QY 5174 CATACGCTGAAACCTTACATCTGACCGGCGCAACCTTCTTACCTTGGTGGCATATCAAGC 5233
 DB 5162 CATACGCTGAAACCTTACATCTGACCGGCGCAACCTTCTTACCTTGGTGGCATATCAAGC 5221
 QY 5234 CCAAAATGAGGTCTCTCACACCCCATTAATTAATCATATGAGGATGATGCGC 5293
 DB 5222 TCAAAATGAGGTCTCTCACACCCCATTAATTAATCATATGAGGATGATGCGC 5281
 QY 5294 TGAACCTGAGGTCTCTCACACCCCATTAATTAATCATATGAGGATGATGCGC 5353
 DB 5282 TGAACCTGAGGTCTCTCACACCCCATTAATTAATCATATGAGGATGATGCGC 5341
 QY 5354 CGCATATCTCCCTGACAGCAGGAGTGTGTCTATTTGGGCGAGATCATCTTCTCCGGGAA 5413
 DB 5342 CGCGTACTCCCTGACAGCAGGAGTGTGTCTATTTGGGCGAGATCATCTTCTCCGGGAG 5401
 QY 5414 GCGAGCTGTCTTCCCGACAGGAGTGTCTCTTACAGGAGTTCGATGAGATGAGAGTGT 5473

PR 11-APR-2001; 2001US-0282965P.
 XX (PTCT-) PTC THERAPEUTICS INC.
 PA Rando R, Weich E;
 XX WPI; 2003-075561/07.
 DR
 XX Identifying a test compound that binds to a target RNA molecule for
 PT treating or preventing amyloidosis, hemophilia, cancer, gigantism,
 PT diabetes, by contacting a detectably labeled target RNA molecule with a
 PT library of test compounds.
 XX Example; Page 64-69; 152pp; English.
 XX
 CC The invention relates to a method for identifying a test compound that
 CC binds to a target RNA molecule, which comprises contacting a detectably
 CC labelled target RNA molecule with a library of test compounds under
 CC conditions that permit direct binding of the labelled target RNA to a
 CC member of the library of test compounds so that a detectably labeled
 CC target RNA: test compound complex is formed. The method is useful for
 CC screening libraries of compounds for those that are selectively bind to a
 CC pre-selected target RNA. The compounds are useful for inhibiting the
 CC formation of a specific bound RNA: host cell factor complexes in vivo.
 CC They are also useful for treating or preventing diseases associated with
 CC overproduction or decreased protein function, such as amyloidosis,
 CC haemophilia, Alzheimer's disease, atherosclerosis, cancer, gigantism,
 CC dwarfism, hypothyroidism, hyperthyroidism, autoimmune disorders, aging,
 CC inflammation, cystic fibrosis, diabetes, obesity, neurodegenerative
 CC disorders, Parkinson's disease or infections (bacterial, viral, fungal).
 CC The invention is also used in gene therapy. The present sequence is
 CC Hepatitis C virus (HCV) DNA. This sequence is used to illustrate the
 CC method of the invention
 XX
 SQ Sequence 9413 BP; 1886 A; 2816 C; 2694 G; 2017 T; 0 U; 0 Other;
 Query Match 86.6%; Score 8311.8; DB 7; Length 9413;
 Best Local Similarity 92.7%; Pred. No. 0;
 Matches 8724; Conservative 0; Mismatches 687; Indels 0; Gaps 0;
 QY 14 TGGGGGGGACATCCACCATGATCACTCCCTGTGAGGAACACTGTCTTCAGGAGAA 73
 DB 2 TGGGGGGGACATCCACCATGATCACTCCCTGTGAGGAACACTGTCTTCAGGAGAA 61
 QY 74 AGCGTCTAGCATGGCGTTAGTGTAGTGTGTCGAGCCCTCCAGGACCCCTCCCGGG 133
 DB 62 AGCGTCTAGCATGGCGTTAGTGTAGTGTGTCGAGCCCTCCAGGACCCCTCCCGGG 121
 QY 134 AGAGCCATAGTGTCTGCGGAAACCGGTGAGTACACCGGAATTGCCAGGACGCGGTCC 193
 DB 122 AGAGCCATAGTGTCTGCGGAAACCGGTGAGTACACCGGAATTGCCAGGACGCGGTCC 181
 QY 194 TTTCTTGGATCAACCGCTCAATGCTGAGATTTGGGCGTGGCCCGGAGACTGCTAG 253
 DB 182 TTTCTTGGATCAACCGCTCAATGCTGAGATTTGGGCGTGGCCCGGAGACTGCTAG 241
 QY 254 CCGAGTAGTCTTGGGTCCGAAAGGCGTGTGTACTGCTGTAGTGGTGTCTTCCGAGTG 313
 DB 242 CCGAGTAGTCTTGGGTCCGAAAGGCGTGTGTACTGCTGTAGTGGTGTCTTCCGAGTG 301
 QY 314 CCGCGGAGGTCTGTTAGACCGGTGCACCATGAGCAGCAATCTTAAACCTCAAGAAAAAC 373
 DB 302 CCGCGGAGGTCTGTTAGACCGGTGCATCATGAGCAGCAATCTTAAACCTCAAGAAAAAC 361
 QY 374 CAACCTTAAACCAACCGCGCCACAGGAGCTCAAGTTCGCGGCGGTGGTCAATCGT 433
 DB 362 CAACCTTAAACCAACCGCGCCACAGGAGCTTAAAGTTCGCGGCGGTGGTCAATCGT 421
 QY 434 TGGTGGAGTTTACCTGTTGCGCGACGCGGCGCCAGGTGTGGTGTGCGCGGCTAGGAA 493
 DB 422 TGGTGGAGTTTACCTGTTGCGCGACGCGGCGCCAGGTGTGGTGTGCGCGGCTAGGAA 481
 QY 494 GGCTTCCGAGGCTGCGCAACCTCGTGGAAAGGCGCAACCTATCCAAAAGGCTGCGCGACC 553

DB 482 GACTTCCGAGCGGTGCAACCTCGTGGAAAGGCGCAACCTATCCCAAGGCTGCGCGCC 541
 QY 554 CGAGGCGAGGGCTGGGCTCAGCCCGGGTACCTTTGGCCCTCTATGCAANTGAGGCT 613
 DB 542 CGAGGCTAGGACCTGGGCTCAGCCCGGGTACCTTTGGCCCTCTATGCAANTGAGGCT 601
 QY 614 GGGGTGGGAGGATGGCTCCTGTCAACCCCGGGCTCCCGGCTAGTTGGGGGCGCCACGGA 673
 DB 602 GGGGTGGGAGATGGCTCCTGTCAACCCCGTGGCTCTGGGCTAGTTGGGGGCGCCACAGA 661
 QY 674 CCGCGGCGTAGGTGCGGTAATCTGGGTAAAGTGTATGCAATCCCTTATGCGGCTTGGC 733
 DB 662 CCGCGGCGTAGGTGCGGTAATCTGGGTAAAGTGTATGCAATCCCTTATGCGGCTTGGC 721
 QY 734 CGATCTCATGGGGTACATTCGCTCGGCGCCCTAGTGGGGCGCTGCCAGGCGCTT 793
 DB 722 CGACCTCATGGGGTACATTCGCTCGGCGCCCTAGTGGGGCGCTGCCAGGCGCTT 781
 QY 794 GGCACACGGTGTCCGGGTCTCGAGGAGCGGCTGAATATGCAACAGGGAATTCGCCGG 853
 DB 782 GGCACATGGTGTCCGGGTCTCGAGGAGCGGCTGAATATGCAACAGGGAATTCGCCGG 841
 QY 854 TTGCTCTTCTATCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 913
 DB 842 TTGCTCTTCTATCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901
 QY 914 TTATGAGTGGCGGACAGTGTCCGGGATATACCATGTCAGACGAGTGTCTTCACTCAAG 973
 DB 902 TTACAGGTTGGCGGACAGTGTCCGGGATATACCATGTCAGACGAGTGTCTTCACTCAAG 961
 QY 974 CATGTGTATGAGGAGCGGACGCTCATCATGATATCTCCCGGTGCTGCTGCTGCTGCTGCT 1033
 DB 962 TATGTGTATGAGGAGCGGACGCTCATCATGATATCTCCCGGTGCTGCTGCTGCTGCTGCT 1021
 QY 1034 GGAGGTAAACAGTCTCCGTTGCTGGTACGCTCATCTCCACGCTCGCGGCGAGGAATGC 1093
 DB 1022 GGAGGTAAATTTCTCCGTTGCTGGTACGCTCATCTCCACGCTCGCGGCGAGGAACAG 1081
 QY 1094 CAGCGTCCCACTACAGCAATACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1153
 DB 1082 CAGCATCCCACTACAGCAATACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1141
 QY 1154 CTGCTCCGTTATGCTGCTGGGGATCTCTGCGGATCTATTTTCTGCTGCTGCTGCTGCTGCT 1213
 DB 1142 CTGTTCCGTTATGCTGCTGGGATCTCTGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1201
 QY 1214 CACTTCTCGCTCGCGGATGAGACGCTGAGGACTGCACTGCTCAATCTATCTGCTGCTGCTGCT 1273
 DB 1202 CACTTCTCACTCGCGGATGAGACGCTGAGGACTGCACTGCTCAATCTATCTGCTGCTGCTGCT 1261
 QY 1274 CCATGTATCAGGTCAACGCTGCTGGGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1333
 DB 1262 CCAGGTATCAGGTCAACGCTGCTGGGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1321
 QY 1334 CTTAGTGGTGTGCGAGTCTTCCGATCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1393
 DB 1322 CTTAGTGGTGTGCGAGTCTTCCGATCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1381
 QY 1394 CCACTGGGAGTCTGCGGGCTTGCCTACTATTTCCATGCTGGGAACTGGGCTAAGGT 1453
 DB 1382 CCACTGGGAGTCTGCGGGCTTGCCTACTATTTCCATGCTGGGAACTGGGCTAAGGT 1441
 QY 1454 TCTGATGTTGGGCTACTCTTTTGGCGGCTGAGCGGGGAGACCCACACGACCGGGAGGCT 1513
 DB 1442 CTTGAT 1501
 QY 1514 GCGCGGCCACACACCTCCGGGTTCAGTCCCTTTTCTCATCTGGGGCTCTCAGAAAAAT 1573
 DB 1502 GGTAGCTCCAGACCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1561
 QY 1574 CCACTGTTGTAATCAACCGGAGTGGGACATCAACAGGACTGCTTAAATGCAATGA 1633

Db 1562 CCAACTCGTGAACACAAACGGCAGCTGGCACATCAACAGCAGCCGCTCTGAATTCGAATGA 1621
QY 1634 CTCCTCCAAACATGGGTTCTTTGCGCGCTGTTTACGGACACAAAGTTCAACTGCTCCGG 1693
Db 1622 CTCCTCCAAACATGGGTTCTTTGCGCGCTGTTTACGGACACAAAGTTCAACTGCTCCGG 1681
QY 1694 GTCCCGGAGCGCATGGCCAGCTGCGCGCCCAATTGACTGCTTCCGCCAGGGGTGGGGCCC 1753
Db 1682 GTGCCAGAGCGCATGCTAGCTGCGCGCCCAATGATGAGCTTGGCTCAGGGGTGGGGTCC 1741
QY 1754 CATCACCTATACAGCTAACAGCTCGGATCAGAGGCTTATGCTGGCAATTAACGGCC 1813
Db 1742 CATCACTCATATGCTCGAGAGCTCGGACAGAGGCAATATGCTGGCACTAAGCGCC 1801
QY 1814 TCGACCGTGGGTTCTGTTACCGGCTCGCAGGTGTTGGTCCAGTGTATGTTTCAACCC 1873
Db 1802 TCGACCGTGGGATCGTGGCTCGCAGGTGTTGGTCCAGTGTATGTTTCACTCC 1861
QY 1874 AAGCCCTGTTGTGGTGGGACCAACCGATCGTTCCGCTGCTCCCTACGTATAGCTGGGGGA 1933
Db 1862 GAGCCCTGTTGTAGTGGGACGACCGATCGTTCCGCGCTCCTACGTATAGCTGGGGGA 1921
QY 1934 GAATGAGACAGAGTGTCTCTCAACACACAGCGTCCGCCACAGAGCAACTGCTCCG 1993
Db 1922 GAATGAGACAGAGTGTCTCTCAACACACAGCGCGCCCTCAAGGCAACTGCTTGG 1981
QY 1994 CTGTACATGATGAATAGTACTGGGTTCACTAAGACGTGGGAGGTCCCGGCTGAACAT 2053
Db 1982 GTGACGTGATGAACAGCACTGGTTCAACAGACGTGGCGGGCCCTCCGTGCAACAT 2041
QY 2054 CGGGGGTGGTAAACCGCACTGTGTCGCCCAACGACTGCTTCGGGAAGACACCCCGA 2113
Db 2042 CGGGGGTGGTAAACCAACCTGTGTTCTGCCCAACGAAATGCTTCGGGAAGACACCCCGA 2101
QY 2114 GGCCTCTTACAAATAGTGGCTCGGGCCCTGGTTGACACCTAGTGCCTAGTACTA 2173
Db 2102 GGCCTCTTACAAAGTGTGGCTCGGGCCCTGGTTGACACCCAGTGTGATGTACTA 2161
QY 2174 CCCATACAGCTTTGGCACTACCCCTGCACTCAATTTTTCCATCTTTAAGGTAGGAT 2233
Db 2162 CCCATACAGCTCTGGCACTACCCCTGCACTGTTACTTTACGTTCTTAAAGTCAAGAT 2221
QY 2234 GTATGGGGGGGTGGAGACAGAGCTCAATGCGGATGCAATGAGCTCGAGAGAGCG 2293
Db 2222 GTATGGGGGGGTGGAGACAGAGCTCAATGCTGCAATGCAATGCACTCGAGAGAGCG 2281
QY 2294 CTGTAACTTGGAGACAGGATAGTTCAGAACTCAGCCGCTGCTGTGTCTAACAAGA 2353
Db 2282 CTGTGACTTGGAGACAGGATAGTTCAGAACTCAGCCGCTGCTGTGTCTAACAAGA 2341
QY 2354 GTGGAGATAGTCCCTGTGCTTTCACACCCCTACCGGCTTTATCCACTGGTTGATCCA 2413
Db 2342 GTGGAGATAGTCCCTGTGCTTTCACACCCCTACCGGCTTTATCCACTGGTTGATCCA 2401
QY 2414 TCTTCCATCAGAACTGTTGACGTGCAATACCTGTAGCGTGTAGGTCAGCGTTGTCTC 2473
Db 2402 TCTTCCACCGAACTGTTGACGTGCAATACCTGTAGCGTGTAGGTCAGCGTTGTCTC 2461
QY 2474 CTGTGCAATCAATGGAGTACATCTGTTGCTTTTCTCTCTCTGACACCGCGCGT 2533
Db 2462 CTGTGCAATCAATGGAGTATATCTGTTGCTTTTCTCTCTCTGACACCGCGCGT 2521
QY 2534 GTGTGCTGTTGGATGATGCTGTATAGCCCAAGCTGAGCGCCCTTAGAGAACTT 2593
Db 2522 CTGTGCTGTTGGATGATGCTGTATAGCCCAAGCTGAGCGCCCTTAGAGAACTT 2581
QY 2594 GGTGTCTCAATGCGCGCTCCGTGGCGAGGCAATGATGATGATGATGATGATGATGAT 2653
Db 2582 GGTGTCTCAATGCGCGCTCCGTGGCGAGGCAATGATGATGATGATGATGATGATGAT 2641
QY 2654 CTGTGCTGCGCTGTTGATATTAAGGCGAGGCTGGCTCTGCGGCGGCGGCTATGCTTTTA 2713
Db 2642 CTGTGCTGCGCTGTTGATATTAAGGCGAGGCTGGCTCTGCGGCGGCGGCTATGCTTTTA 2701

QY 2714 TGGCGTATGGCGCTGCTCCTGCTCCTACTGAGCTTACCAACAGAGCTTACGCTTGA 2773
Db 2702 TGGCGTATGGCGCTGCTCCTGCTCCTACTGAGCTTACCAACAGAGCTTATGCCATGA 2761
QY 2774 CGGGAGATGGCTGTCATCGTGGGGGTGCGGTTCTTGTAGGTCTGGTATTTTACGCTT 2833
Db 2762 CGGAGATGGCTGTCATCGTGGAGGCGCGTTTGTAGGTCTGGTACTCTTACGCTT 2821
QY 2834 GTACCACTATCAAAAGTGTCTCACTAGGCTCATATATGTTGTTGTTTACATCTTTATCAC 2893
Db 2822 GTACCACTATCAAAAGTGTCTCCTAGGTCTATATGTTGTTGTTTACATCTTTATCAC 2881
QY 2894 CAGAGCGAGGCGCATGCAAGTGTGGTCCCGCTTCAAGCTTCGGGAGGCGCGA 2953
Db 2882 CAGAGCGAGGCGCATGCAAGTGTGGTCCCGCTTCAAGTTCGGGAGGCGCGA 2941
QY 2954 TGCCATCATCTCCTCAAGTGTGGTTCATCAGAGTAAATTTTGTGATCATCAACAACT 3013
Db 2942 TGCCATCATCTCCTTACATGCGGCTTCAATCAGAGCTAATCTTTGATCATCAACAACT 3001
QY 3014 CTTGCTCGCATACTGCGCGCTCATAGTGTCTCAGGCTGSCATAACGAGAGTGCCTGA 3073
Db 3002 CTTGCTCGCATACTGCGGCTCATAGTGTCTCAGGCTGSCATAACGAGAGTGCCTGA 3061
QY 3074 CTTGCTCGGCTCAAGGCTCATCTGTCATGATGTTAGTGGAAAGTCCCGGGG 3133
Db 3062 CTTGCTCGGCTCAGGCTCATCTGTCATGATGTTAGTGGAGAGTCCCTGGAG 3121
QY 3134 TCTATATGTCCTAAATGGTCTTCAATGAAGTGGGCGCTGACAGAGTACGTAGCTTTATAA 3193
Db 3122 CCACTATGTCCTAAATGGCTTCAATGAAGTGGCGCTGACAGTACGTAGCTATATGA 3181
QY 3194 CCACTATGTCCTAAATGGCTTCAATGAAGTGGCGCTGACAGAGTACGTAGCTTTATAA 3253
Db 3182 CCACTATGTCCTAAATGGCTTCAATGAAGTGGCGCTGACAGAGTACGTAGCTTTATAA 3241
QY 3254 AGAGCCGCTGCTCTTCCGCTACAGAGTCACTGCGGAGCAGACACCG 3313
Db 3242 AGAGCCGCTGCTCTTCTGACATGAGAGTAACTCATCACTGCGGAGCAGACACCG 3301
QY 3314 TGGCTGAGGACATCATCTTGGGCTTACCGCTTCCGCGCCGAGGAGAGATAT 3373
Db 3302 GCGCTGAGGACATCATCTGCGGCTTACCGCTTCCGCGCCGAGGAGAGATAT 3361
QY 3374 TTTGAGCCGCTGATAGTTCGAGAGGCAAGGGTGGGACTCTTGGCGCCCATCACGC 3433
Db 3362 TCTAGACCCGCGCTATGTTTGGAGAGAGGGGTGGGCTCTTGGCTTATCACGC 3421
QY 3434 CTACTCCCAACAAACGCGGGGTACTTTGGTTCATCATCACTAGCCTCACAGGCGGGA 3493
Db 3422 CTATTCCCAACAAACGCGGGGCTTGGTGGTGTATCATCACTAGCCTCACAGGCGGGA 3481
QY 3494 CAGAGACAGGTCGAGAGGGAGTTCAGTGTCTTACCGCAACACATCTTTCCTGGC 3553
Db 3482 CAGAGACAGGTCGAGTGGGAGTTCAGTGTCTTCCACCGCAACGCAATCTTTCCTGGC 3541
QY 3554 GACCTGATCAACGCGCTGCTGACTGTCTACCATGCGCTGCTCGAAGACCTAGC 3613
Db 3542 GACCTGCTCAATGGCTGTGTTGGAGCCGTCTACATGTTGCGGCTCGAAGACCTAGC 3601
QY 3614 CGGTCCAAAGTCCAAATCACCCAAATGTACCAATGTAGACCTTGGACCTCGTGGCTG 3673
Db 3602 CGGCGCGGAGGTTCCAAATCACCCAAATGTACCAATGTAGACCTTGGACCTCGTGGCTG 3661
QY 3674 GAGAGGCGCGCGGCGGCTCATGACACATGACAGTGTGAGCTTGGACCTTGA 3733
Db 3662 GCGCGCGCGCGGCGGCTCATGACACCTGACACCTGACAGCTTGA 3721
QY 3734 CTTGCTCAGAGACATGCTGATGTCATTCCGCTGCGCGGCGGAGGAGCAGAGGAGGAG 3793
Db 3722 CTTGCTCAGAGACATGCTGATGTCATTCCGCTGCGCGGCGGAGGAGCAGAGGAGGAG 3781

5942	CGAGATGCCCTCCACTGAGGANTCTGGTTAAATTTACTCCCTGCCATCTTTCTCTGGCGC	6001
6014	CCTGTCCTCGGGGTCGTGTGCGCAGCAATACTGCGTCGGCACGTGGCCCGCGAGAGGG	6073
6002	CCTGTTGTCGGGGTCGTGTGCGCAGCAATACTGCGTCGGCACGTGGCCCGCGAGAGGG	6061
6074	GGCTGTGCAGTGGATGAACCGGCTCATAGCGTTGCTTCGGGGGTAAACACGTCCTCC	6133
6062	GGCTGTGCAGTGGATGAACCGGCTCATAGCGTTGCTTCGGGGGTAAACACGTCCTCC	6121
6134	TACGCACCTATGTGCTGAGAGCGACGGCTGACACGTCATCTCAGATCCCTCTCTAGCCT	6193
6122	CACGCACCTATGTGCTGAGAGCGACGGCGCGCGCGTGTACTCAGATCCCTCTCCAGCCT	6181
6194	TACCATCACTCAACTGCTGGAAGCGGCTCCACGCTGGATTAATGAGGACTCTCTACGCC	6253
6182	TACCATCACTCAGTTGCTGGAAGCGCTTCATCAGTGGATTAATGAGGACTCTCTCAGGCC	6241
6234	ATGCTCCGGCTCGTGGCTAAAGGATGTTTGGATTTGGATATGCAACGTTGTGACTGACTT	6313
6242	TTGTTCCGGCTCGTGGCTTAAGGATGTTTGGACTGGATATGCACGGTGTAGTGACTT	6301
6314	CAAGACTGGCTCAGTCCAACTCTCTCCCGGTTAOCGGAGTCCCTTTCTGTCTCATG	6373
6302	CAAGACTGGCTCCAGTCCAAAGCTCTCTCCCGGTTAOCGGAGTCCCTTTCTGTCTCATG	6361
6374	CCAAOCGGGTCAAGGCGAGTCTGCGGGGGAGCGGCATCATGAACCACTCCGCCATG	6433
6362	CCAAOCGGGTCAAGGCGAGTCTGCGGGGGGATGGCATCATGAACCACTCCGCCATG	6421
6434	CGGAGCTCAGATTCGCCCGACATGTCAAAAACGGTTCCATGAGGATCTGATGGCCCTAGAAC	6493
6422	TGGAGCTCAGATTCACCGACATGTCAAAAATGGCTCCATGAGGATGTGTGGGCCAAAAAC	6481
6494	CTGCAGCAACACGTGGCAGCAAGCTTCCCATCAACGATACACCGAGACCTTCGCAC	6553
6482	CTGCAGCAACACGTGGCATGGAAATTTCCCATCAACGATACACCGAGACCTTCGCAC	6541
6554	ACCTTCCC CGCGCCCAACTATTCCAGGGCGCTATGCGGGTGCTCTCAGAGGATACGT	6613
6542	GCCCTCCCAGCGCCGAATATTCCAGGGCGCTGTGGCGGGTGCTCTGAGAGATACGT	6601
6614	GGAGGTTACCGCTGTGGGGNTTTCCACTAGCTGACGGGCATGACCACTGACAAACGTAAA	6673
6602	GGAGGTTACCGGGTGGGGNTTTCCACTAGCTGACGGGCATGACCACTGACAAACGTGAA	6661
6674	GTGCCCATGCCAGGTTCCGGCCCCGAAATTTCTTCAACGAGGTGATGGAGTCGGGTTGCA	6733
6662	ATGCCCATGCCAGGTTCCAGCCCCGAAATTTTTCACGAGGTGGATGGAGTACGGTTGCA	6721
6734	CAGGTACGCTCCGGCTTGCAACCTCTTTCTACGGGAGGACGTACGTTCCAGGTCGGGCT	6793
6722	CAGGTATGCTCCAGTGTGCAACCTCTTCTTCAAGAGAGGTGCTGATTTCCAGGTCGGGCT	6781
6794	CAACCAATATCTGTTCGGGTTCGACGCTCCCATGCGAGCCCGCAACCGACGTTAACAGTGCT	6853
6782	CAACCAATATCTGTTCGGGTTCAGCTCCCATGTGACCCCGAACCCGATGTGCGAGTGCT	6841
6854	TACTTCATGCTACCGATCCCTCCCACTTACAGCAGAGACGGCTTAAGCTTAGGCTGGC	6913
6842	CACPTCCATGCTACCGACCCCTCTCATATTTACAGCAGACACGGCCCAAGGTAGGCTGGC	6901
6914	TAGAGGCTTCCCCCTCTTTAGCCAGCTCATAGCTAGCCAGTGTGCTCGGCTCTCTTT	6973
6902	CAGGGGTCTCCCCCTCTCTTGGCCAGCTCTTTAGCTAGCCAGTGTGCTCGGCTCTCTTT	6961
6974	GAAGGCACATGCACTACCACTATGCTCCCGGACGCTGACCTCATTCAGGCGCAACCT	7033
6962	GAAGGCACATGTACTACCCATCATGCTCCCGGACGCTGACCTCATTCAGGCGCAACCT	7021
7034	CTTTGTGGCGCAGAGATGGCGGAAACATCATCTCGCTGAGATCAGAGAAATAGGTAGT	7093
7022	CCCTGTGCGGACAGAGATGGCGGGAAACATCACCCGTGTGAGTCAAGAAATAGGTGCT	7081

QY	7094	AATTTGGACTCTTTTCGAACGGCTTCACGGGAGGGGGATAGAGGGAGATATCCGTGCG	7153
DB	7082	AATCTGTGACTCTTTTCGATCCGATTTCGGCGGTGGAGGATGAGAGGAAATATATCCGTCCC	7141
QY	7154	GGCGGAGATCTGTGAAAAATCCAGGAAGTTCCCTCAGCGCTTCGCAATATGGGCAAGCC	7213
DB	7142	GGCGGAGATCTGTGAAAAACCGAGGAATTCCTCCCGAGCGTTGCCCATATGGGCAAGCC	7201
QY	7214	GGACTAAATCTCTCACTGCTAGAGTCTCTGGAAGACCCGCACTACGTCCCTCCGTTGT	7273
DB	7202	GGATTAAACCTCTCACTGTAGAGTCTCTGGAAGACCCGCACTACGTCCCTCCGTTGT	7261
QY	7274	ACCGGATGCCATTGGCACTCACCAAGGCTCTCCATACCACTCCACGGAAGAAG	7333
DB	7262	ACCGGATGCCATTGGCACTCACCAAGGCCCCCAATACCACTCCACGGAAGAAG	7321
QY	7334	GACGGTTGCTCTGACAGAAATCCATGTGTCTTCTGCTTGGCGGAGCTGCCACTAAGAC	7393
DB	7322	GACGGTTGCTCTGACAGAGTCCACCGTGTCTTCTGCTTGGCGGAGCTCGTACTAAGAC	7381
QY	7394	CTTCGTAGCTCCGGATCTGTGCGCGTTGATAGGGCAAGCGGACCGCTTCTGTACCT	7453
DB	7382	CTTTGGAGCTCCGGGTCTGTGCGCGTTGACAGCGGACCGCGACTGGCCCTCCCGATCA	7441
QY	7454	GCCCTCCGACGACGGTGCACAAAGATCCGACGTTTGAGTCTGTCTCTCATGCCCCCT	7513
DB	7442	GCCCTCCGACGACGGCGACAAAGGATCCGACGTTTGAGTCTGTCTCTCATGCCCCCT	7501
QY	7514	TGAAGGGAGCGGGGGACCCCGATCTCAGCAGCGGTCTTGTCTCACCGTAGTGAGGA	7573
DB	7502	CGAGGAGAGCGAGGGACCCCGACCTCACGCAAGGTCTTGTCTTACCGTGAGCGGGA	7561
QY	7574	GCTTAGTAGAGTGTGCTGTCTGTCTCAATGCTCTATCTGTGACAGGCGCCGTGATCA	7633
DB	7562	AGCTGTGTAGGACGTGTGTCTGTCTCAATGTCTTATATGACAGAGGTGCTTGTATCAC	7621
QY	7634	GCCATCGCTCCGAGAGAAATGAAGCTGCCCATCAACCGCTTGAGCAACTCTTTGCTGCG	7693
DB	7622	GCCATCGCTCCGAGAGAGCAAGTTGCCCATCAATCCGTTGAGCAACTCTTTGCTGCG	7681
QY	7694	TCACCAACAATGGTCTACGCGCAACATCCCGACGCGAAGCTCGCGAGAGAGGT	7753
DB	7682	TCACCAAGATGGTCTACTCCACAATCTCGCAGCGCAAGTCTCGCGAGAGAGGT	7741
QY	7754	CACCTTTGACAGATTGCAAGTCTCTGATGATCAATACCGGAGCTACTCAAGGAGATGAA	7813
DB	7742	CACCTTTGACAGACTGCAAGTCTCTGACGACCACTACCGGAGCTCTCAAGGAGATGAA	7801
QY	7814	GGGAGGGCTCCACAGTTAAGCTTAGCTTCTATCTATAGAGAGGCTTCGAGCTGAC	7873
DB	7802	GGGAGGGCTCCACAGTTAAGCTTAGCTTCTATCTATAGAGAGGCTTCGAAACTGAC	7861
QY	7874	GCCCCACATTCGGGCAAAATCCAAAATTTGGCTATGGGCAAGGAGCTCGGAACCTATC	7933
DB	7862	GCCCCACATTCGGGCAAAATCCAAAATTTGGCTACGGGGCGAGGAGCTCGGAGCTATC	7921
QY	7934	CAGAGGGCGTTAAACCAATCCGCTCCGTGTGGGAGGACTTGTGGAAGACACTGAAAC	7993
DB	7922	CAGAGGGCGGTCAACCAATCCGCTCCGTGTGGGAGGACTTGTGGAAGACACTGAAAC	7981
QY	7994	ACCAATTGACACCAACCATATGCAAAAAGTGAAGTTTCTGGGTCCACACGAGAGGG	8053
DB	7982	ACCAATTGATACCAACCATATGCAAAAATGAGGTTTCTGGGTCCACACGAGAGGG	8041
QY	8054	AGGCGGCAAGCCAGCTCGCTTATCTGATTCGCAGACCTGGGAGTTGTGTATGGAGAA	8113
DB	8042	AGGCGGCAAGCCAGCTCGCTTATCTGATTCGCAGACCTGGGAGTTGTGTATGGAGAA	8101
QY	8114	GATGGCCCTTACGAGTGTCTGACCCCTTCTCAGGCCGTGATGGGCTCTCTATACGG	8173
DB	8102	GATGGCCCTTACGAGTGTCTCACCCTTCTCAGGCCGTGATGGGCTCTCTATACGG	8161

QY 8174 ATTTCATTAATCTCCCTCCCAAGCAGCGGGTCGAGTTCTCTGGTGAATACCTGGAAATCAAGAA 8233
 Db 8162 ATTCAGTACTCTCTGGGAGCGGGTCGAGTTCTCTGGTGAATACCTGGAAATCAAGAA 8221
 QY 8234 ATGCCCTATGGGCTTCTCATATGACACCCGCTGTTTTGACTCAACGGTCACTGAGTGGA 8293
 Db 8222 ATGCCCTATGGGCTTCTCATATGACACCCGCTGTTTTGACTCAACGGTCACTGAGAAATGA 8281
 QY 8294 CATTCGGTGTTCAGGAGTCAATTTTACCAATGTTGAGACTTGGCCCGCCGAGGCCACAGAGGC 8353
 Db 8282 CATCCGTACTGAGGAATCAATTTTACCAATGTTGAGACTTGGCCCGCCGAGGCCACAGAGGC 8341
 QY 8354 CATAAAGTTCGCTCAACAGAGCGGCTTTTATCATCGGGGGTCCCTGTACTAACTCAAAAGGGCA 8413
 Db 8342 CATAAAGTTCGCTCAACAGAGCGGCTTTTATCATCGGGGGTCCCTGTACTAACTCAAAAGGGCA 8401
 QY 8414 GAACCTGGGTATTCGCGGTGCGCGCAAGTGGGCTGTCTGACGACTAGCTGCGGTATATAC 8473
 Db 8402 GAACCTGGGTATTCGCGGTGCGCGCAAGTGGGCTGTCTGACGACTAGCTGCGGTATATAC 8461
 QY 8474 CCTCACATGTACTTGAAGGCCACTGACGCTGTGAGCTGCAAGCTTCCAGGACTGCAC 8533
 Db 8462 CCTCACATGTACTTGAAGGCCACTGACGCTGTGAGCTGCAAGCTTCCAGGACTGCAC 8521
 QY 8534 GATCTCTGAAACGAGAGACCTTGTGTTATCTGTAAAGCGCGGAAACCCAGGAGGA 8593
 Db 8522 GATCTCTGAAACGAGAGACCTTGTGTTATCTGTAAAGCGCGGAAACCCAGGAGGA 8581
 QY 8594 TGGCGGCGCTAGAGGCTTCAAGGAGGCTATGACTAGGTATTCGGGCGCCCGCGGGA 8653
 Db 8582 TGGCGGCGCTAGAGGCTTCAAGGAGGCTATGACTAGGTATTCGGGCGCCCGCGGGA 8641
 QY 8654 TCCGCCGCCAACAGAAATGAGGCTGAGCTGATTAACATCATGTTCTCCCAATGTGCTAGT 8713
 Db 8642 CCGGCCCCCAACAGAAATGAGGCTGAGCTGATTAACATCATGTTCTCCCAATGTGCTAGT 8701
 QY 8714 CGGCAAGATGCACTTGGCAAAAGGTTATACCTCAACCGGTGACCCACCCCGCT 8773
 Db 8702 CGGCAAGATGCACTTGGCAAAAGGTTATACCTCAACCGGTGACCCACCCCGCT 8761
 QY 8774 TGCAGGGCTGCTGGGAGACAGCTAGACACACTTCCATCACTTCTGGCTAGGCAATAT 8833
 Db 8762 CGCAGGGCTGCTGGGAGACAGCTAGACACACTTCCATCACTTCTGGCTAGGCAATAT 8821
 QY 8834 CATCATGTATGCGGCCACCTTATCGGCAAGGATGATTCGATGACTCACTTTTCTCAT 8893
 Db 8822 CATCATGTATGCGGCCACCTTATCGGCAAGGATGATTCGATGACTCACTTTTCTCAT 8881
 QY 8894 CTTTCTAGCTCAAGAGCAACTTGAAGAGCCCTGATTTGATGATCTACCGGCTGCTGCTA 8953
 Db 8882 CTTTCTAGCTCAAGAGCAACTTGAAGAGCCCTGATTTGATGATCTACCGGCTGCTGCTA 8941
 QY 8954 CTCATTGAGCCACTGACCTTACTCTGATCACTTGAACGACTCCATGCTTACGCAAT 9013
 Db 8942 CTCATTGAGCCACTGACCTTACTCTGATCACTTGAACGACTCCATGCTTACGCAAT 9001
 QY 9014 TACACTTCAAGTACTTCTCAGGTGAGATCAATAGGTTGCTTTCATGCCCTCAGGAACT 9073
 Db 9002 TTAATCCACAGTACTTCTCAGGTGAGATCAATAGGTTGCTTTCATGCCCTCAGGAACT 9061
 QY 9074 TGGGGTACCACTTGGGAACTTGGAGACATCGGGCCAGAGTGTCCGCGCTAAGTACT 9133
 Db 9062 TGGGGTACCACTTGGGAACTTGGAGACATCGGGCCAGAGTGTCCGCGCTAAGTACT 9121
 QY 9134 GTCCAGGGGGGGGGCGGCGCACTTGGGAGATACCTTTTAACTGGGAGTAAAGAC 9193
 Db 9122 GTCCAGGGGGGGGGCGGCGCACTTGGGAGATACCTTTTAACTGGGAGTAAAGAC 9181
 QY 9194 CAAGCTTAAATCTCACTCAATCCCGGCGCGCTCCAGCTGACCTTGTCTGGCTGCTTGGT 9253
 Db 9182 CAAGCTTAAATCTCACTCAATCCCGGCGCGCTCCAGCTGACCTTGTCTGGCTGCTTGGT 9241
 QY 9254 CGCTGGTTACAGCGGGGAGACATATATCAAGGCTGTCTCTGTCGCCGACCCCGCTGGTT 9313

Db 9242 TGTGTGTACACGGGGAGACATATACACCTGTCTCGTCCGACCCCGTTGGTT 9301
 QY 9314 TCCGTTGTGCTACTCTCTACTTCTGTAGGGTAGGCAATTAAGTCTCCGACCGATG 9373
 Db 9302 CATGTTGTGCTACTCTCTACTTCTGTAGGGTAGGCAATTAAGTCTCCGACCGATG 9361
 QY 9374 AACGGGAGGTAACCACTCCAGGCTTAAAGCCATTTCCCTGTTTTTTTTTTT 9424
 Db 9362 AACGGGAGGTAACCACTCCAGGCTTAAAGCCATTTCCCTGTTTTTTTTTTT 9412

RESULT 6
 AAL53723
 ID AAL53723 standard; DNA; 9413 BP.
 XX
 AC AAL53723;
 XX
 DT 27-OCT-2003 (revised)
 DT 07-FEB-2003 (first entry)
 XX
 DE Hepatitis C Virus DNA sequence SEQ ID No 17.
 XX
 KW Target RNA; target RNA:support-attached test compound; flow cytometry;
 KW mass spectrometry; high-throughput screening; ds.
 XX
 OS Hepatitis C virus; Virus.
 XX
 FN WO200283837-A1.
 XX
 PD 24-OCT-2002.
 XX
 PF 11-APR-2002; 2002WO-US011758.
 XX
 PR 11-APR-2001; 2001US-0282966P.
 XX
 PA (PTCT-) PTC THERAPEUTICS INC.
 XX
 PI Almstead NG;
 XX
 DR WPI; 2003-075534/07.
 XX
 PT Identifying a test compound that binds to a target RNA molecule by
 separating the detectably labeled target RNA:support-attached test
 compound complex from uncomplexed target RNA molecules and test compounds
 by flow cytometry.
 XX
 PS Example; Page 55-59; 131pp; English.
 XX
 CC The invention relates to a novel method for identifying a test compound
 that binds to a target RNA molecule comprising separating the detectably
 labeled target RNA:support-attached test compound complex from
 uncomplexed target RNA molecules and test compounds. The separating
 process is carried out by flow cytometry and determining a structure of
 the type of test compound of the RNA:support-attached test compound
 complex by mass spectrometry. The method is useful for high-throughput
 screening of libraries of compounds to identify pharmaceutical leads.
 CC This polynucleotide sequence represents a DNA sequence related to the
 detecting method of the invention. (Updated on 27-OCT-2003 to standardise
 OS field)
 CC
 SQ Sequence 9413 BP; 1887 A; 2815 C; 2694 G; 2017 T; 0 U; 0 Other;
 Query Match 86.6%; Score 8310.2; DB 7; Length 9413;
 Best Local Similarity 92.7%; Pred. No. 0;
 Matches 8723; Conservative 0; Mismatches 688; Indels 0; Gaps 0;

QY 14 TGGGGGGGCACTCCACCATGAATCACTCCCTGTGAGGAATCACTGTCTTCACGAGAA 73
 Db 2 TGGGGGGGCACTCCACCATGAATCACTCCCTGTGAGGAATCACTGTCTTCACGAGAA 61
 QY 74 AGGCTTAGCCATGGCGCTTAGTATGAGTTCGTCAGCCTCCAGGACCCCTCCCGG 133

Db	62	AGCGTCTAGCCATGGCGCTTAGTAGAGTGTTGTGACGCTCCAGGACCCCGCCCTCCGGG	121
Qy	134	AGAGCCATAGTGGTCTCGGAAACGGGTGAGTACACCGGAATGCCAGAGACGACCGGTCC	193
Db	122	AGAGCCATAGTGGTCTCGGAAACGGGTGAGTACACCGGAATGCCAGAGACGACCGGTCC	181
Qy	194	TTCTCTGGATCAACCCGCTCAATGCCCTGGAGATTGGCGGTGCCCGCGAGACTGCTAG	253
Db	182	TTTCTTGAGTCAACCCGCTCAATGCCCTGGAGATTGGCGGTGCCCGCGAGACTGCTAG	241
Qy	254	CCGAGTAGTGTGGGTCCGAAAGGCCCTTGTTGGTACTCCCTGATAGGTCCTTCCGAGTG	313
Db	242	CCGAGTAGTGTGGGTCCGAAAGGCCCTTGTTGGTACTCCCTGATAGGTCCTTCCGAGTG	301
Qy	314	CCCCGGAGGCTCTGTTAGACGCTGTGCACATGACGACGAATCCTAAACCTCAAGAAAAAC	373
Db	302	CCCCGGAGGCTCTGTTAGACGCTGTGCATCATGAGCACAAATCCTAAACCTCAAGAAAAAC	361
Qy	374	CAAAAGTAAACCAACCGCCGCCACAGGACGTCAGGTTCCCGGGCGGTGGTCAGATCGT	433
Db	362	CAAAAGTAAACCAACCGCCGCCACAGGACGTTAAGTTCCCGGGCGGTGGTCAGATCGT	421
Qy	434	TGGTGGAGTTTACCTGTTTGGCGACAGGGGCCAGGTTGGGTGTGGCGCGACTAGGAA	493
Db	422	TGGTGGAGTTTACCTGTTTGGCGACAGGGGCCAGGTTGGGTGTGGCGCGACTAGGAA	481
Qy	494	GGCTTCCGAGCGGTGCGAACTCGTGGAGGCGCAACCTATCCAAAGAGCTCGCGAACC	553
Db	482	GACTTCCGAGCGGTGCGAACTCGTGGAGGCGCAACCTATCCCAAGGCTTCGCGGCC	541
Qy	554	CGAGGCGAGGCTGGGCTCAGCCCGGGTACCCTTGGCCCTCTATGGCAATGAGGCGCT	613
Db	542	CGAGGCTAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCCTCTATGGCAACGAGGCTAT	601
Qy	614	GGGTGGCGAGGATGCTCTCTGCACCCCGGCTCCGGGCTAGTTGGGGCCCCACGGA	673
Db	602	GGGTGGCGAGGATGCTCTCTGCACCCCGGCTCTCGGCTAGTTGGGGCCCCACAGA	661
Qy	674	CCCCCGGCTAGGTGCGGTAACTGGGTAAAGGTCATGATACCTTATATCGCGCTTCGC	733
Db	662	CCCCCGGCTAGGTGCGGTAACTGGGTAAAGGTCATGATACCTTATATCGCGCTTCGC	721
Qy	734	CGATCATAGGGTACATTCGGCTCTGGGGCCCCCTAGGGGGCGCTGCCAGGCGCTT	793
Db	722	CGACTCATGGGGTACATTCGGCTCTGGGGCCCCCTAGGGGGCGCTGCCAGGCGCTT	781
Qy	794	GGCACACGCTGTCGGGTCTCTGGAGGACGGCGTGAACCTAGCAACAGGGAACATTGCCCGG	853
Db	782	GGCACATGGTGTCCGGTCTCTGGAGGACGGCGTGAACCTAGCAACAGGGAATCTGCCCGG	841
Qy	854	TTGCTCTTTCTCTATCTTCTCTTGGCTCTGCTGCTCTGTTTGACATCCAGCTTCCGC	913
Db	842	TTGCTCTTTCTCTATCTTCTCTTGGCTCTGCTGCTCTGTTTGACATCCAGCTTCCGC	901
Qy	914	TTATGAAGTCCGCAACGTTGTCGGGATATACCATGTCAAGAACAGCTGCTCCAACTCAAG	973
Db	902	TTACGAGTTCGGCAACGTTGTCGGGATATACCATGTCAAGAACAGCTGCTCCAACTCAAG	961
Qy	974	CATTGTGTATGAGCGACGGGACGTGATCATGCACTATCCCGGGTGGCTGCTGTGTCTCA	1033
Db	962	TATTGTGTATGAGCGACGGGACATGATCATGCAACCCCGGGTGGCTGCTGTGTCTCA	1021
Qy	1034	GGAGGTAACAGCTCCCGTTGCTGGGTAGCGCTACCTCCACGCTCCGCGCAGGATGC	1093
Db	1022	GGAGGTAATTTCTCCGTTGCTGGGTAGCGCTACCTCCACGCTCCGCGCAGGATGC	1081
Qy	1094	CAGGCTCCCCACTACGAAATACACGCCACGTCGATCTGCTCGTTGGGACGGCTGCTTT	1153
Db	1082	CAGCATCCCCACCACGAATACAGCCACGTCGATCTGCTCGTTGGGACGGCTGCTTT	1141
Qy	1154	CTGCTCCGCTATGTACGTGGGGGATCTCTCGGGAATCTATTTTCTGCTCTCCAGCTGTT	1213
Db	1142	CTGTTCCGCTATGTACGTGGGGATCTCTCGGATCCGTTTCTGCTCTCCAGCTGTT	1201

Db 4442 AGAGATTCCCTTCTATGGCAAAACCCATCCCAATGGAGCCATCAAGGGGGAGGCAATCT 4501
QY 4514 CATTTCTGCTCATCTCAAGAAAGAAATGTGACGAGCTCGCGCAAAAGCTGACAGGCTCGG 4573
Db 4502 CATCTTCTGCCATTCAGAGAGAGGTGTGACGAGCTCGCGCAAAAGCTGACAGGCTCGG 4561
QY 4574 ACTGAACGCTGTAGCATATTAACCGGGCCCTTGATGTGCTGCTCATCCGCTATCCGAGA 4633
Db 4562 ACTCAATGCTGTAGCGTATTAACCGGGCTCTCGATGTGCTCATCCGACTAGCGAGA 4621
QY 4634 CGTCTGTCTGCTGCAACAGAGAGCTCTAATGAGCGGTTTACCGGCGATTTGACATCAGT 4693
Db 4622 CGTCTGTCTGCTGCAACAGAGAGCTCTAATGAGCGGTTTACCGGCGATTTGACATCAGT 4681
QY 4694 GATCGATGCAATATCATGTGTCAACAGAGAGCTGAGCTTTCAGCTTGGATCCCACTTCAC 4753
Db 4682 GATCGACTGCAACACATGTGTCAACAGAGAGCTGAGCTTTCAGCTTGGATCCCACTTCAC 4741
QY 4754 CATTTAGAGAGAGAGCTGTGCCCCAAGAGCGGTGTGCGGTGTGCGAATCGGCGAGGTAGAAC 4813
Db 4742 CATTTAGAGAGAGAGCTGTGCCCCAAGAGCGGTGTGCGGTGTGCGAATCGGCGAGGTAGAAC 4801
QY 4814 TGGCAGGGGTAGAGTGGGATCTACAGGTTTGTGACTCCAGAGAGAGCGGCTCGGCGAT 4873
Db 4802 TGGCAGGGGTAGAGTGGGATCTACAGGTTTGTGACTCCAGAGAGAGCGGCTCGGCGAT 4861
QY 4874 GTTCGATTTCTCGTCTGTGTGAGTGCTATGACGGGGCTGTGCTGTGATGAGCTTCAC 4933
Db 4862 GTTCGACTCTCGTCTGTGTGAGTGCTATGACGGGGCTGTGCTGTGATGAGCTTCAC 4921
QY 4934 GCCCGGTGAGAGCTCGGTGTAGGTTGGGCTTACCTTAATATACACCGAGGTGTCGCTGTG 4993
Db 4922 GCCCGGTGAGAGCTCGGTGTAGGTTGGGCTTACCTTAATATACACCGAGGTGTCGCTGTG 4981
QY 4994 CCAGGACCAATCTGGAGTTCTGGAGAGCGTCTTCAAGGCTTCAACACATAGATGCCA 5053
Db 4982 CCAGGACCAATCTGGAGTTCTGGAGAGCGTCTTCAAGGCTTCAACACATAGATGCCA 5041
QY 5054 CTTCTGTCTCCAGCTAAACAGGAGAGAGCACTTTCTTACCTGTGTCATATCAAGC 5113
Db 5042 CTTCTGTCTCCAGCTAAACAGGAGAGAGCACTTTCTTACCTGTGTCATATCAAGC 5101
QY 5114 TACAGTGTGCGCAGGCTCAAGCTCACCTCCATCTGTGGAGCAATGTGGAGTGTCT 5173
Db 5102 CACAGTGTGCGCAGGCTCAAGCTCACCTCCATCTGTGGAGCAATGTGGAGTGTCT 5161
QY 5174 CATACGCTGAAACCTACATGACGCGGCCAAACACCCCTCTGTATAGGCTAGGAGCGGT 5233
Db 5162 CATACGCTGAAACCTACATGACGCGGCCAAACACCCCTCTGTATAGGCTAGGAGCGGT 5221
QY 5234 CCAAAATGAGGTATCTTCAACACCCCAATCTAATATATCATATGGCATGTGTCGCG 5293
Db 5222 TCAAAATGAGGTATCTTCAACACCCCAATCTAATATATCATATGGCATGTGTCGCG 5281
QY 5294 TGACCTGTGAGGTCTGCTACTAGCACTGGGTGTGTAGGGGAGTCTTTCAGCTTGGC 5353
Db 5282 TGACCTGTGAGGTCTGCTACTAGCACTGGGTGTGTAGGGGAGTCTTTCAGCTTGGC 5341
QY 5354 CGCATACTGCTGACGACAGGAGTGTGTGCTATGTTGGGAGGATCATCTTGTCCGGGAA 5413
Db 5342 CGCGTACTGCTGACGACAGGAGTGTGTGCTATGTTGGGAGGATCATCTTGTCCGGGAG 5401
QY 5414 GCCAGCTGTCTGTTCCGACAGGAGTCTTCTACAGGAGTTCATGAGATGGAAGTG 5473
Db 5402 GCCAGCTGTCTTATCCGACAGGAGTCTTCTACAGGAGTTCATGAGATGGAAGTG 5461
QY 5474 TGCCTCAAACTTCTTACATCGAGAGGAGTGTGCTGCGGAGCAATTCAGGAA 5533
Db 5462 TGCTTCAACTCTTCTTACATCGAGAGGAGTGTGCTGCGGAGCAATTCAGGAA 5521
QY 5534 GGGCTCGGTTGTGCAACCGGCCAACAGAGCGGAGGCTGCTGCTCCCGTGTGGA 5593
Db 5522 GGGCTCGGTTGTGCAACCGGCCAACAGAGCGGAGGCTGCTGCTCCCGTGTGGA 5581

QY 5594 GTCCAAAGTGGGAGCCCTTGACACCTTCTGGCGAAGCACATGTGGAATTTTCATCAGCG 5653
Db 5582 GTCCAAAGTGGGAGCCCTTGAGGCTTCTTGGGCGAACAACATGTGGAATTTTCATCAGCG 5641
QY 5654 AATACAGTACTAGCAGGCTTATCCACTCTGCTGGGAAACCCCGGATAGCATCATTTGAT 5713
Db 5642 GATACAGTACTTGGCAGGCTTATCCACTCTGCTGGGAAACCCCGGATAGCATCATTTGAT 5701
QY 5714 GGCATTTACAGCTTCTATACATAGCCCTCACACCCCAAAACACCTTCTGTTTAAACAT 5773
Db 5702 GGCATTTACAGCTTCTATACACAGCCCTCACACCCCAAAATACCTTCTGTTTAAACAT 5761
QY 5774 CTTGGGGGATGGGTGGCTGCCCAACTCTCTCCAGAGCTGCGTCAGCTTTCGTGG 5833
Db 5762 CTTGGGGGATGGGTGGCTGCCCAACTCTCTCCAGAGCTGCGTCAGCTTTCGTGG 5821
QY 5834 CGCCGATCGCCGAGAGGCTGTGTTGGAGAGATAGSCCTTGGGAAGTGTCTGTGAGCAT 5893
Db 5822 CGCCGATCGCCGAGAGGCTGTGTTGGAGAGATAGSCCTTGGGAAGTGTCTGTGAGCAT 5881
QY 5894 CTTGGCGGCTATGCGGAGAGGCTAGCCGCGCACTCTGTGGCTTTAAAGTCAACGAGCG 5953
Db 5882 TCTGGCGGCTATGCGGAGAGGCTGTGTTGGAGAGATAGSCCTTGGGAAGTGTCTGTGAGCAT 5941
QY 5954 CGAGTGTCCCTCAACGAGGACCTGTCTCAACTTACTTCCCTGCCATCTCTCTCTGGTGC 6013
Db 5942 CGAGATGTCCCTCAACTGAGGATCTGTGTTAAATTTACTTCCCTGCCATCTCTCTCTGGTGC 6001
QY 6014 CTTGTCTCGGGGTGTGTGCGAGCATACTCTGTGCGACGTGTGGACGTGGCCCGGAGAGGG 6073
Db 6002 CTTGTGTGTGGGGTGTGTGCGAGCAATACTGTGTGCGACGTGTGGCCCGGAGAGGG 6061
QY 6074 GCTGTGAGTGCATCAACCGGCTCATAGCTGTGCTTGGCGGGGTAAACAGCTCTCCCG 6133
Db 6062 GCTGTGAGTGCATCAACCGGCTCATAGCTGTGCTTGGCGGGGTAAACAGCTCTCCCG 6121
QY 6134 TAGCAGTATGTCTGTGAGAGGAGCTGTGACAGAGCTGTGCACTCAGATCTCTCTGAGCT 6193
Db 6122 CACGCACTATGTGTCCGAGAGGAGCGCGCGGCGGTGTACTCAGATCTCTCTCAGCT 6181
QY 6194 TACCATCTCACTGCTGAAGGGCTCCACAGTGGATTAATAGGAGTGTCTTACGCG 6253
Db 6182 TACCATCTCACTGCTGAAGAGGCTTCACTGAGGAGTCACTGAGGATTAATAGGAGTGTCTTACGCG 6241
QY 6254 ATGTCTCCGCTCGTGGCTAAGGATGTTTGGATGATATGACAGGCTGTGAGTCACTT 6313
Db 6242 TTGTTCCGCTCGTGGCTAAGGATGTTTGGAGTGGATATGACAGGCTGTGAGTCACTT 6301
QY 6314 CAAAGCTGCTTCCAGTCCAAACTCTGTGCGGGGTACCGGAGTCCCTTCTCTCATG 6373
Db 6302 CAAAGCTTGGCTCCAGTCCAACTCTGTGCGGGGTACCGGAGTCCCTTCTCTCATG 6361
QY 6374 CAAACGCGGCTCAAGGAGTCTGCGGGGGAGCGGCACTATGCAAAACCACTGCGCCATG 6433
Db 6362 CAAACGCGGCTCAAGGAGTCTGCGGGGGAGTGGCATCATGCAAAACCACTGCGCCATG 6421
QY 6434 CGGAGCAGATCGCGGAGCATGTCAAAACCGTTTCCATGAGGATGCTAGGGGCTTAGAAC 6493
Db 6422 TGGAGCAGATCAACCGGACATGTCAAAATGGCTTCCATGAGGATGTTGTTGGGCCAAAAAC 6481
QY 6494 CTGACAGACAGTGGCAGGAACTGTCCCACTCAACGATACACAGGAGCTTGGAC 6553
Db 6482 CTGACAGAAACAGTGGCAGTGGAACTTCCCACTCAACGATACACAGGAGCTTGGAC 6541
QY 6554 ACCCTTCCCGGCGCCCACTATTTCCAGGGCGCTTATGGCGGGTGTGCTGAGGAGTACGT 6613
Db 6542 GACCTTCCCGAGCGCCAACTTTCAGGGCGCTGTGCGGGGTGGCTGTGAGGAGTACGT 6601
QY 6614 GAGGTTACGCTGTGGGGATTTCCACTCTGTGAGGAGTGTGACCACTGACCAAGTAAA 6673
Db 6602 GAGGTTACGCGGGTGGGGATTTCCACTCTGTGAGGAGTGTGACCACTGACCAAGTAAA 6661

QY 6674 GTGCCATGCGAGTTCGGGCCCCGGAATTTCTTCAACGAGGTGATGGAGTGGGTTGCA 6733
DB 6662 ATGCCCATGCCAGTTTCAGCCCTGATTTTTCACGAGGTGATGAGTTGCA 6721
QY 6734 CAGTACGCTCCGGGTCGAAACCTCTTCTACGGGAGAGCTCAGTTTCCAGTTCGGGCT 6793
DB 6722 CAGTATGCTCCAGTGTGCAACCTCTCTACGAGAGGAGTCTGATTTCCAGGTCGGCT 6781
QY 6794 CAACCAATACTTGTTCGGTTCGAGTCCCATGAGCCGAGCCGCAACCGGACGCTACAGTCT 6853
DB 6782 CAACCAATACTTGTTCGGTTCGAGTCCCATGAGCCGAGCCGCAACCGGACGCTACAGTCT 6841
QY 6854 TACTTCATGCTACCCGATCCCTCCACATTTACAGCAGAGAGCGCTAAGCTTAGGCTGGC 6913
DB 6842 CACTTCATGCTACCCGATCCCTCTCATATTTACAGCAGAGAGCGCTAAGCTTAGGCTGGC 6901
QY 6914 TAGAGGCTCTCCCTCTTTAGCCAGCTCATCAGCTAGCCAGTGTCTCGGCCCTTTT 6973
DB 6902 CAGGGGCTCTCCCTCTTTAGCCAGCTCATCAGCTAGCCAGTGTCTCGGCCCTTTT 6961
QY 6974 GAAGGCCACATGCACTACCAACATGACTCCCGGAGCGTGAACCTCATCGAGGCGCAACCT 7033
DB 6962 GAAGGCCACATGCACTACCAACATGACTCCCGGAGCGTGAACCTCATCGAGGCGCAACCT 7021
QY 7034 CTTGTGGCGCAGAGATGGCGGAAACATCACTCGCGTGGAGTCAAGATTAAGTGTGT 7093
DB 7022 CTTGTGGCGCAGAGATGGCGGAAACATCACTCGCGTGGAGTCAAGATTAAGTGTGT 7081
QY 7094 AATCTCGACTCTTTCGAAACGCTTCACGGGAGGGGATGAGAGGAGATATCCCTGCG 7153
DB 7082 AATCTCGACTCTTTCGAAACGCTTCACGGGAGGGGATGAGAGGAGATATCCCTGCG 7141
QY 7154 GGGGAGATCCTTGGAATAACGAAATTCCTTCAGCGTTCGCCATATGGGCAACGCC 7213
DB 7142 GGGGAGATCCTTGGAATAACGAAATTCCTTCAGCGTTCGCCATATGGGCAACGCC 7201
QY 7214 GGACTACATCTTCCATCTGTAGAGTCTTGGAGGACCGGACTACGTCCTTCGGTGGT 7273
DB 7202 GGAATTAACAACCTTCCATCTGTAGAGTCTTGGAGGACCGGACTACGTCCTTCGGTGGT 7261
QY 7274 ACACGATGCCATGCGACCTACAGGCTTCCTCAATACCACTCCACGAGAGAGAG 7333
DB 7262 ACACGATGCCATGCGACCTACAGGCTTCCTCAATACCACTCCACGAGAGAGAG 7321
QY 7334 GACGGTTGCTTGCACAGAAATCCATGTCTTCTGCTTCGGGAGCTCGCCACTAAGAC 7393
DB 7322 GACGGTTGCTTGCACAGAGTCCACCGGTCTTCTGCTTCGGGAGCTCGCTACTAGAC 7381
QY 7394 CTTTCGGTACCTCCGATCGTGGCGGTGATAGCGGACCGGACCGCCCTTCCTGACCT 7453
DB 7382 CTTTCGGTACCTCCGATCGTGGCGGTGATAGCGGACCGGACCGCCCTTCCTGACCT 7441
QY 7454 GGCCTCCGACGAGTGAAGAAGTCCGAGTTCGCTGCTGCTTCTTCCTCCATGCCCCCT 7513
DB 7442 GGCCTCCGACGAGTGAAGAAGTTCGCTGCTGCTTCTTCCTCCATGCCCCCT 7501
QY 7514 TGAAGGGAGCCGGGGACCCCGATCTCAGCGAGGCTTGTGCTACCGTGAAGGA 7573
DB 7502 CGAGGGAGAGCCAGGGAGCCCGACCTCAGCGAGGCTTGTGCTACCGTGAAGGA 7561
QY 7574 GCTAGTGAAGATGCTGCTGCTGCTCAATGTCTTATAGTGAACGAGGCGCCCTGATCAC 7633
DB 7562 AGCTGGTGAAGATGCTGCTGCTCAATGTCTTATAGTGAACGAGTGCCTGATCAC 7621
QY 7634 GCCATGCTGCGGAGGAAAGTGAAGTCCCATCAACCCCTGAGCAACTCTTTGCTGG 7693
DB 7622 GCCATGCTGCGGAGGAGAGTGAAGTCCCATCAATCCCTGAGCAACTCTTTGCTGG 7681
QY 7694 TCACCAACAATGCTTACGCCCAACAATCCCGAGGCGCAAGCTCCCGGAGAGAGGT 7753
DB 7682 TCACCAACAATGCTTACGCCCAACAATCCCGAGGCGCAAGCTCCCGGAGAGAGGT 7741
QY 7754 CACTTTGACAGATTGCAAGTCTTGGATGATCATTAACGGGAGCGTACTCAAGGAGATGAA 7813

DB 7742 CACTTTGACAGACTGCAAGTCTTGCAGACCTACCGGACGCTGCTCAAGAGATGAA 7801
QY 7814 GGGGAAGGGCTCACAAGTTAAGCTAAGCTTCTATCTATAGAGAGGCTCGAAGCTGAC 7873
DB 7802 GGGGAAGGGCTCACAAGTTAAGCTAAGCTTCTATCTATAGAGAGGCTCGAAGCTGAA 7861
QY 7874 GCGCCACATTCGGCCAAATCCAAATTTGGCTATGGGCAAGAGACGTCGGGACCTATC 7933
DB 7862 GCGCCACATTCGGCCAAATCCAAATTTGGCTATGGGCAAGAGACGTCGGGACCTATC 7921
QY 7934 CAGCAGGGCGTTAACCAATCCGCTCCGTTGGGAGGACTTGTCTGGAAGACACTGAAAC 7993
DB 7922 CAGCAGGGCGCTCAACCAATCCGCTCCGTTGGGAGGACTTGTCTGGAAGACACTGAAAC 7981
QY 7994 ACCAATGTACACACCAATCATATGSCAAAGAGTGGGTTTCTGCGTCCACAGAGAGGG 8053
DB 7982 ACCAATGTATACCAATCATATGSCAAAGAGTGGGTTTCTGCGTCCACAGAGAGGG 8041
QY 8054 AGCCCGCAAGCCAGCTCGCTTATCGTATTTCCAGACCTGGGAGTTCTGTATGCGAGAA 8113
DB 8042 AGCCCGCAAGCCAGCTCGCTTATCGTATTTCCAGACCTGGGTTTCTGTATGCGAGAA 8101
QY 8114 GATGCGCTTTACGAGTGTCTCCACCTTCTCAGGCGGTGATGGGTCTCTCATACGG 8173
DB 8102 GATGCGCTTTACGAGTGTCTCCACCTTCTCAGGCGGTGATGGGTCTCTCATACGG 8161
QY 8174 ATTTCAATACTCCCCCAAGCAGCGGCTGAGTTCCTGCTGGAATACCTGGAATCAAGAA 8233
DB 8162 ATTTCAATACTCCCCCAAGCAGCGGCTGAGTTCCTGCTGGAATACCTGGAATCAAGAA 8221
QY 8234 ATGCTTATGGGCTTCTCATATGACACCGCTGTTTGTGATCAACCGTCACTGAGAGTGA 8293
DB 8222 ATGCTTATGGGCTTCTCATATGACACCGCTGCTTGTGATCAACCGTCACTGAGAGTGA 8281
QY 8294 CATTCGTTGTAGGAGTCAATTTACCAATGTTGTGACTTGGCCCCCGGAGCCACAGAGC 8353
DB 8282 CATTCGTTGTAGGAGTCAATTTACCAATGTTGTGACTTGGCCCCCGGAGCCAGAGC 8341
QY 8354 CATTAAGGTGCTCACAAGAGGGCTTTACATCGGGGTCCTGACTAATCAAAAGGCA 8413
DB 8342 CATTAAGGTGCTCACAAGAGGGCTTTATGTCGGGGTCCCTGACTAATTCGAAGGGGCA 8401
QY 8414 GAATCGGTTATTCGGCGGTGCGGCAAGTGGCGTCTGACGACTAGCTGCGGTATATC 8473
DB 8402 GAATCGGTTATTCGGCGGTGCGGCAAGTGGCGTCTGACGACTAGCTGCGGCAACAC 8461
QY 8474 CTTCAATGTTACTTGAAGSCCACTGACGCTGTGAGCTGGAAGCTCCAGAGCTGCAC 8533
DB 8462 CTTCAATGTTACTTGAAGSCCACTGCGGCTTCTGAGCTGCAAAAGCTCCAGAGCTGCAC 8521
QY 8534 GATGCTGCTGAACGAGAGCGACCTTCTGCTTATCTGTGAAGCGCGGAAACCGAGAGGA 8593
DB 8522 GATGCTGCTGAACGAGAGCGACCTTCTGCTTATCTGTGAGAGTGGGGAAACCGAGAGGA 8581
QY 8594 TCGCGGGCTTACGAGCTTTCACGAGGCTATGACTAGGTATTCGCGCCCCCGGAGGA 8653
DB 8582 TCGCGGGCTTACGAGCTTTCACGAGGCTATGACTAGGTATTCGCGCCCCCGGAGGA 8641
QY 8654 TCGCGCCCAACAGATATGACCTGGAGCTGATAACATCATGTTCTTCAATGTGTGCT 8713
DB 8642 CCGCGCCCAACAGATATGACCTTGGAGCTGATAACATCATGTTCTTCAATGTGTGCT 8701
QY 8714 CCGGCAAGTATGCTGCGGAGAGGCTATACCTGACCTGACCCGAGCCACCCCT 8773
DB 8702 CCGGCAAGTATGCTGCGGAGAGGCTATACCTGACCTGACCCGAGCCACCCCT 8761
QY 8774 TGCACGGGCTGCTGGGAGAGAGCTAGACACATCTCAATCAACTTTGGTATGGCAATAT 8833
DB 8762 CGCAGGGCTGCTGGGAGAGAGCTAGACACACTCAGTCAACTCTTGGCTAGGCAATAT 8821
QY 8834 CATCATGTTATGCGCCCAACCTATGGCAAGGATGATTTCTGATGACTCACTTTTCTCCAT 8893

```

Dd 8822 CATCATGTATGGCCCAACCCATATGGGCGAGGATGATCTCATGACTCATTTCTCTCAT 8881
QY 8894 CCTCTAGCTCAAGAGCAACTTGAAGAGCCCTGGATTGTGAGATCTAACGGGCTTGCTA 8953
Dd 8882 CCTCTAGCTCAGAGCAACTTGAAGAGCCCTGGATTGTGAGATCTAACGGGCTTGCTA 8941
QY 8954 CTCGATTGAGCGCTTGAACCTACCTCAGATCATGATCATGAAAGCACTCATGCTTTAGCGCAT 9013
Dd 8942 CTCGATTGAGCGCTTGAACCTACCTCAGATCATGATCATGAAAGCACTCATGCTTTAGCGCAT 9001
QY 9014 TACACTCCACAGTTACTCTCAGGTGAGATCAATAGGCTGGCTTCATGCTCCAGAACT 9073
Dd 9002 TTCACTCCACAGTTACTCTCAGGTGAGATCAATAGGCTGGCTTCATGCTCCAGAACT 9061
QY 9074 TGGGTTACCACTCTGCAACCTTGGAGACATCGGGCCAGAGTGTCCGCGCTAAGCTACT 9133
Dd 9062 TGGGTTACCACTCTGCAACCTTGGAGACATCGGGCCAGAGTGTCCGCGCTAAGCTACT 9121
QY 9134 GTCCAGGGGGAGGGCCGCACTTGTGGAGATACCTTTTAACCTGGGCGATAGGAC 9193
Dd 9122 GTCCAGGGGGAGGGCCGCACTTGTGGAGATACCTTTTAACCTGGGCGATAGGAC 9181
QY 9194 CAAGCTTAAACTCACTCCAACTCCCGCGCGGTCCAGCTGAGCTGTGCTGCTGCTGCT 9253
Dd 9182 CAAGCTTAAACTCACTCCAACTCCCGCGCGGTCCAGCTGAGCTGTGCTGCTGCTGCT 9241
QY 9254 CGCTGTTACAGGGGGAGAGATATATACAGCCTGTCTCGTCCCGAGCCCGCTGCTT 9313
Dd 9242 TGCTGTTACAGGGGGAGAGATATATACAGCCTGTCTCGTCCCGAGCCCGCTGCTT 9301
QY 9314 TCCTGTTGCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9373
Dd 9302 CATGTTGCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9361
QY 9374 AACGGGAGCTAACCACTCCAGGCGCTTAAGCCATTTCTCTCTCTCTCTCTCTCTCT 9424
Dd 9362 AACGGGAGCTAACCACTCCAGGCGCTTAAGCCATTTCTCTCTCTCTCTCTCTCTCT 9412

RESULT 7
AAT03960
ID AAT03960 standard; cDNA; 9413 BP.
XX
AC AAT03960;
XX
DT 16-OCT-2003 (revised)
DT 14-NOV-1996 (first entry)
XX
XX Partial HCV non-structural coding region.
XX
XX protease; hepatitis C virus; screening; inhibitor; proteolytic;
XX identification; cleavage; ss.
XX
XX Hepatitis C virus; Virus.
XX
XX Key Location/Qualifiers
XX CDS 330..932
XX /tag= a
XX mat_peptide 3021..4028
XX /tag= b
XX /note= "partial proteinase; see AAT03958"
XX mat_peptide 3303..6050
XX /tag= c
XX /note= "partial proteinase; see AAT03959"
XX
XX JF07184648-A.
XX
XX 25-JUL-1995.
XX
XX 05-FEB-1993; 93JP-00018854.
XX
XX 07-FEB-1992; 92JP-00022657.
XX 18-SEP-1992; 92JP-0024240.

```

```

PR 04-DEC-1992; 92JP-00325303.
XX
XX (KAEN/) KAENNO K.
PA (SUMQ) SUMITOMO METAL IND LTD.
XX (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
XX
XX WPI; 1995-287962/38.
DR DR
XX P-PSDB; AAR82694.
XX
XX An HCV proteinase active substance - which has activity as an anti-HCV
XX agent and can be used to screen for proteinase inhibitors.
XX
XX Disclosure; Page 39-48; 52pp; Japanese.
XX
XX The present sequence is a partial Hepatitis C Virus (HCV) sequence from
XX the non-structural region. Partial proteinase encoding sequences (AAT03958
XX - 59) are contained within this sequence. The proteinases can be used as
XX anti-HCV agents. They can also be used to screen cpds. for their ability
XX to inhibit their proteolytic activity. In this way proteinase inhibitors
XX can be identified. (Updated on 16-OCT-2003 to standardise OS field)
XX
XX Sequence 9413 BP; 1887 A; 2816 C; 2693 G; 2017 T; 0 U; 0 Other;
XX
XX Query Match 86.6%; Score 8308.6; DB 2; Length 9413;
XX Best Local Similarity 92.7%; Pred. No. 0;
XX Matches 8722; Conservative 0; Mismatches 689; Indels 0; Gaps 0;
XX
QY 14 TGGGGCGGACACTCCACCATGAATCACTCCCTGTGAGAACTACTGTCTTCACGAGAA 73
Dd 2 TGGGGCGGACACTCCACCATGAATCACTCCCTGTGAGAACTACTGTCTTCACGAGAA 61
QY 74 AGCGTTAGCCATGCGGTTAGTATGAGTGTCTGAGCCTCCAGGACCCCTCCCGGG 133
Dd 62 AGCGTTAGCCATGCGGTTAGTATGAGTGTCTGAGCCTCCAGGACCCCTCCCGGG 121
QY 134 AGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGGAATTCGAGACGACCGGGTCC 193
Dd 122 AGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGGAATTCGAGACGACCGGGTCC 181
QY 194 TTTCTTGATCAACCGGCTCAATGCTGGAGATTTGGGGGTGCCCCCGGAGACTGCTAG 253
Dd 182 TTTCTTGATCAACCGGCTCAATGCTGGAGATTTGGGGGTGCCCCCGGAGACTGCTAG 241
QY 254 CCGAGTGTGTTGGGTTCGCGAAGGCTTGTGTACTGTCTGATAGGTTGCTTGGAGTG 313
Dd 242 CCGAGTGTGTTGGGTTCGCGAAGGCTTGTGTACTGTCTGATAGGTTGCTTGGAGTG 301
QY 314 CCGCGGAGGTCTCGTAGCCGTGACCATGAGACGAAATCTTAACCTCAAGAAAGAAC 373
Dd 302 CCGCGGAGGTCTCGTAGCCGTGACCATGAGACGAAATCTTAACCTCAAGAAAGAAC 361
QY 374 CAAAGCTAACACCAACCGCGGCCACAGGACGTCAAGTTCCCGGGGTGCTGAGATGCT 433
Dd 362 CAAAGCTAACACCAACCGCGGCCACAGGACGTCAAGTTCCCGGGGTGCTGAGATGCT 421
QY 434 TGGTGGAGTTTACCTGTTCCCGGAGGGGCCCGAGGTGGGTGTCGCGGAGCTAGGAA 493
Dd 422 TGGTGGAGTTTACCTGTTCCCGGAGGGGCCCGAGGTGGGTGTCGCGGAGCTAGGAA 481
QY 494 GGCCTCCGAGCGGTCCGAACCTCGTGGAGGGGCAACCTATCCCAAGGCTCGCGGACC 553
Dd 482 GACTTCCGAGCGGTCCGAACCTCGTGGAGGGGCAACCTATCCCAAGGCTCGCGGACC 541
QY 554 CGAGGCGAGGGCTGGGCTCAGCCCGGCTACCTTTGGCCCTCTATGGCAATGAGGGCT 613
Dd 542 CGAGGCTAGGACTTGGGCTCAGCCCGGCTACCTTTGGCCCTCTATGGCAAGGCTAT 601
QY 614 GGGGTGGGAGGATGCTCTCTGTCACCGCGGTCTCCGCGCTAGTTGGGGCCCAAGAA 673
Dd 602 GGGGTGGGAGGATGCTCTCTGTCACCGCGGTCTCCGCGCTAGTTGGGGCCCAAGAA 661
QY 674 CCGCGGAGGTGCTGCTGCTTGGGTAAAGTTCATCGATACCTTACCTTACCTGCGTTCG 733

```

Db 662 CCCCCGGCTAGGTGCGGTAAATTTGGGTAAAGTCAATGATACCCCTTACATGCGGCTTCGC 721
Qy 734 CGATCTCATGGGTACATTCGGCTCCTCGGCGCCCCCTAGGGGGCGCTGCCAGGCGCTT 793
Db 722 CGAAGCTCATGGGTACATTCGGCTCCTCGGCGCCCCCTAGGGGGCGCTGCCAGGCGCT 781
Qy 794 GGCACACGGTTCGGGTTCTGGAGGACGGGCTGAATATGCAACAGGGAATCTTCCCGG 853
Db 782 GGCACATGGTTCGGGTTCTGGAGGACGGGCTGAATATGCAACAGGGAATCTTCCCGG 841
Qy 854 TTGCTCTTTCTCATATCTCTCTTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 913
Db 842 TTGCTCTTTCTCATATCTCTCTTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901
Qy 914 TTATGAAGTGGCAAGGTCTCGGGATATACCATGTGACGAACAGCTGCTCCAACTCAAG 973
Db 902 TTACGAGGTGGCAAGGTCTCGGGATATACCATGTGACGAACAGCTGCTCCAACTCAAG 961
Qy 974 CATTTGTATGAGCAGCGGACGTGATCATCATACTCTCCGGGTGCGCTGCTGCTGCTGCT 1033
Db 962 TATTTGTATGAGCAGCGGACATGATCATCATCATACTCTCCGGGTGCGCTGCTGCTGCTGCT 1021
Qy 1034 GGAGGTAAACGCTCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1093
Db 1022 GGAGGTAAATTTCTCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
Qy 1094 CAGCGTCCCCACTACGCAATACGACGCGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1153
Db 1082 CAGCATCCCCACACGCAATACGACGCGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1141
Qy 1154 CTGCTCCGCTATGATACGTTGGGGATCTCTGCGGATCTATTTTCTGCTGCTGCTGCTGCTGCT 1213
Db 1142 CTGTTCCGCTATGATACGTTGGGGATCTCTGCGGATCCGTTTCTGCTGCTGCTGCTGCTGCT 1201
Qy 1214 CACCTTCTCGCTCGCGGATGAGACGATGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1273
Db 1202 CACCTTCTCGCTCGCGGATGAGACGATGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1261
Qy 1274 CCATGTATCAGGTACCGCATGCTTGGGATATGATGATGATGATGATGATGATGATGATGATGAT 1333
Db 1262 CCAGTATCAGGTACCGCATGCTTGGGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1321
Qy 1334 CTTAGTGGTGTGCGAGTGTCCGGATCCCAACAGCTGTGCTGACATGCTGCTGCGGGGCG 1393
Db 1322 CCTAGTGGTATGCGAGCTACTCCGATCCCAACAGCGCTGCTGACATGCTGCTGCGGGGCG 1381
Qy 1394 CCATGGGGTGTCTGCGGGCTTGGCTTACTATTCATGCTAGGAGTGGGCTAAGGT 1453
Db 1382 CCATGGGGTGTCTGCGGGCTTGGCTTACTATTCATGCTAGGAGTGGGCTAAGGT 1441
Qy 1454 TCTGATTTGGCGCTACTCTTTTGGCGGCTGACGGGGAGACCCACACGACGGGGAGGGT 1513
Db 1442 CTTGATTTGATGCTACTCTTTTGGCGGCTGACGGGGAGACCCACACGACGGGGAGAG 1501
Qy 1514 GCGCGGACACACCTCTCGGGTTCAGTCCCTTTTCTCATGCTGGGCGCTCTCAGAAAT 1573
Db 1502 GGTAGCCTCCAGACCCAGAGCCTCGTGTCTGCTGCTCTCAAGGGCCCATCTCAGAAAT 1561
Qy 1574 CCAGCTTGTGAATACCAACGCGAGCTGCGACATCAACAGGAGTCCCTTAATTCATGA 1633
Db 1562 CCAGCTTGTGAATACCAACGCGAGCTGCGACATCAACAGGAGCCTCTGAATTCAGATGA 1621
Qy 1634 CTCCTCCAACTGGGTCTTTTGGCGGCTGTTTTAGGCAACAAATTCCTGCTCCCGG 1693
Db 1622 CTCCTCCAACTGGGTCTTTTGGCGGCTGTTTTAGGCAACAAATTCCTGCTCCCGG 1681
Qy 1694 GTGCGGAGGGGATGCGAGCTCCCGCCCATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1753
Db 1682 GTGCGGAGGGGATGCGAGCTAGCTCCCGCCCATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1741
Qy 1754 CATCACCTATACCTAAGCTACAGCTCGGATCAGAGGCTTATGCTGGGATTAAGCGGCC 1813
Db 1742 CATCACCTATACCTAAGCTACAGCTCGGATCAGAGGCTTATGCTGGGATTAAGCGGCC 1801

Qy 1814 TCGACCGTGTGGTGTGCTACCCCGGTCGAGGTGTGGTGCAGTGTATTTGTTTACCCC 1873
Db 1802 TCGACCGTGTGGGATCGTGTCTGCTGCGAGGTGTGGTGCAGTGTATTTGTTTACCTC 1861
Qy 1874 AAGCCCTGTGTGGTGGGACCAACCGATCGTTCGGGTGTCCTACGTATAGCTGGGGGA 1933
Db 1862 GAGCCCTGTGTGGTGGGACCAACCGATCGTTCGGGTGTCCTACGTATAGCTGGGGGA 1921
Qy 1934 GAATGACACAGACGTGTGCTCTCAACACACGCGTCCGCCCAAGGCAACCTGGTTCGG 1993
Db 1922 GAATGACACAGACGTGTGCTCTCAACACACGCGTCCGCCCAAGGCAACCTGGTTCGG 1981
Qy 1994 CTGTACATGATGAATAGTACTGGTGTCACTAGACGTGGGAGGTCCTCCGCTGTACAT 2053
Db 1982 GTGACGTGATGAACAGCACTGGTGTCAACAGAGTGGGGGCTCCCTGCTGACAT 2041
Qy 2054 CGGGGGGTGGTAAACGCACTTGTATCTGCCCAACGCACTGCTTCCGGAAGCACCCCGA 2113
Db 2042 CGGGGGGTGGTAAACGCACTTGTATCTGCCCAACGCACTGCTTCCGGAAGCACCCCGA 2101
Qy 2114 GGCTACTTACAAAAATGTGGTCTGGGGGCTCTGGTGAACCTAGGTGCTAGTACATA 2173
Db 2102 GGCACCTTACAAAAATGTGGTCTGGGGGCTCTGGTGAACCTAGGTGCTAGTACATA 2161
Qy 2174 CCCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTTCATCTTTAAGGTTAGGAT 2233
Db 2162 CCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTTCATCTTTAAGGTTAGGAT 2221
Qy 2234 GTATGTGGGGGGTGGAGCACAGGCTCAATGCGCATGCAATGGACTCGAGGAGAGCG 2293
Db 2222 GTATGTGGGGGGTGGAGCACAGGCTCAATGCGCATGCAATGGACTCGAGGAGAGCG 2281
Qy 2294 CTGTACTTGGAGGACAGGATAGGTGAGAACTCAGCCCGCTGCTGTCTTCAACAGAG 2353
Db 2282 CTGTACTTGGAGGACAGGATAGGTGAGAACTCAGCCCGCTGCTGTCTTCAACAGAG 2341
Qy 2354 GTGCGAGATACCTGCTGCTTCCACCACTACCGCTTTTATCCACTGTTGATCCA 2413
Db 2342 GTGCGAGATACCTGCTGCTTCCACCACTACCGCTTTTATCCACTGTTGATCCA 2401
Qy 2414 TCTCCATCAAGACATCGTGGACGTGCAATACCTGTACGGGTAGGGTCAAGCTTTGCTC 2473
Db 2402 TCTCCATCAAGACATCGTGGACGTGCAATACCTGTACGGGTAGGGTCAAGCTTTGCTC 2461
Qy 2474 CTTTGAATCAAAATGGAGTATATCCTGTGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2533
Db 2462 CTTTGAATCAAAATGGAGTATATCCTGTGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2521
Qy 2534 GTGTGCTGCTTCTGTGATGATGCTGTAGTCCAGGCTGAGGCGCTTAGAGAACTT 2593
Db 2522 CTGTGCTGCTTCTGTGATGATGCTGTAGTCCAGGCTGAGGCGCTTAGAGAACTT 2581
Qy 2594 GGTGGTCTCAATGCGCGCTCGTGGCGAGCGCATGCTTCTCTCTCTCTCTCTCTCTCTCTCT 2653
Db 2582 GGTGGTCTCAATGCGCGCTCGTGGCGAGCGCATGCTTCTCTCTCTCTCTCTCTCTCTCTCT 2641
Qy 2654 CTTTCTGGCGCTTGTATCAATTAAGGCGAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2713
Db 2642 CTTTCTGGCGCTTGTATCAATTAAGGCGAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2701
Qy 2714 TGGCGTATGGCGCT 2773
Db 2702 TGGCGTATGGCGCT 2761
Qy 2774 CCGGAGATAGGCTGATCTGCGGGGCTCGGTTCTTTGAGGCTGCTGTTCTTTGACCTT 2833
Db 2762 CCGGAGATAGGCTGATCTGCGGGGCTCGGTTCTTTGAGGCTGCTGTTCTTTGACCTT 2821
Qy 2834 GTCACCATACTACAAAGTGTCTCTACTAGGCTCATATGGTGGTTACATATCTTTATCAC 2893
Db 2822 GTCACCATACTACAAAGTGTCT 2881

Qy	2894	CAGAGCGAGCGGCACATGCAAGTGTGGTGTCCCCCCTCAAAGTTTGGGGAGGCGCGGA	2953
Db	2882	CAGAGCGAGCGGCACATGCAAGTGTGGTGTCCCCCCTCAAAGTTTGGGGAGGCGCGGA	2941
Qy	2954	TGCATCATCTCTCTCAAGTGTGGGTTCATCCAGAGTTAAATTTTGCATCACCAAACT	3013
Db	2942	TGCCAICATCTCTTACATGGCGGTCCATCCAGAGCTAACTTTTGACATCACAAACT	3001
Qy	3014	CTCTCTCGCATATCTGGCGCGCTCATGTGTCTCCAGGCTGGCATAACGAGAGGCGGTA	3073
Db	3002	CTTCTCGCCATATCTGGTCCGCTCATGTGTCTCCAGGCTGGCATAACGAGAGGCGGTA	3061
Qy	3074	CTTCTGGCGGCTCAAGGGCTCATTCGTGATGATGTAGTGTGGAAAGTGTGCGGGGG	3133
Db	3062	CTTTGTAGCGCTCAGGGGCTCATTCGTGATGATGTAGTGTGGAAAGTGTGCGTGAGG	3121
Qy	3134	TCATATGTCCAAATGTCTTTCATGAAGTGTGGCGGCTGACAGGTACGTACGTTTATAA	3193
Db	3122	CCATATGTCCAAATGTCTTTCATGAAGTGTGGCGGCTGACAGGTACGTACGTTATATGA	3181
Qy	3194	CCATCTTACCCACTCGCGGACTGGGCGCCACGGGGCTACGAGACCTTGGGTGGCGGT	3253
Db	3182	CCATCTTACTCCATCGCGGANTTGGGCGCCACGGGGCTACGAGACCTTGGGTGGCGGT	3241
Qy	3254	AGAGCCCGTGTCTTCTCCGCCATGTGAGACCAAGGTCACTACTGGGGAGCAGACACCGC	3313
Db	3242	AGAGCCCGTGTCTTCTTGACATGGAGACTAAACTCATCACTCTGGGGGACAGACACCGC	3301
Qy	3314	TGGGTGTGGGACATCATCTTTGGTCTACCGGTCTCCGCCGAAGGGGAGAGAGATATT	3373
Db	3302	GGCGTGTGGGACATCATCTCTGGGTCTACAGGTCTCCGCCGAAGGGGAGAGAGATACT	3361
Qy	3374	TTTGGGACCCGCTCATATGTCTCGAAGGGCAAGGTGTGGACTCTCTTGGCGCCATCAACGC	3433
Db	3362	TCTAGGACCGCCGATATGTTTGGAGAGCAGGGTGTGGGCTCTTGGCGCTATCACCGC	3421
Qy	3434	CTACTTCCAAACAACGGCGGGCTATTTGGTTTGCATCATCACTAGCTCTCAGCGCGGGA	3493
Db	3422	CTATTTCCAAACAACGGCGGGCTCTTGGTGTGTATCATCACTAGCTCTCAGCTCGGGA	3481
Qy	3494	CAAGAACAGGTTCGAAGGGAGGTTCAAGTGGTTTCTACCGCAACAATCTTCTCTGGC	3553
Db	3482	CAAGAACAGGTTCGATGGGAGGTTCAAGTGTCTCTCCACGCAACCAATCTTCTCTGGC	3541
Qy	3554	GACCTGATCAACGGGTGTCTGCACTGTCTACCATGGGCGCTGGCTCGAAGACCTTAGC	3613
Db	3542	GACCTGCTCAATGGCGTGTGTGCAACGCTTACCATGGTCCGGCTCGAAGACCTTGGC	3601
Qy	3614	CGGTCCAAAAGGTCCAATCAACCAATGTACCAATGTAGACTGTGACCTCTGTGGCTG	3673
Db	3602	CGGCCCCAAGGTTCCATCACCCCAATGTACCAATGTAGACAGGACCTCTGTGGCTG	3661
Qy	3674	GCAGGCGCCCCCGGGCGGCTCTCATGACACCATGCAAGTGTGGCAGCTCTGGACCTTTTA	3733
Db	3662	GCAGGCGCCCCCGGGCGGCTCTCATGACACCGTGCACCTGGCGAGCTCTGACCTTTA	3721
Qy	3734	CTTTGGTCAACGAGACATCTCATGTTCATCTCGGTGTGGCGGAGGCGCAGCAGAGGGGAAG	3793
Db	3722	CTTTGGTCAACGAGGATCTCATGTCTCGGTGTGGCGGAGGCGCAGCAGCAGAGGGGAG	3781
Qy	3794	TCTACTCTCCCCAAGGCGGCTCTCTTACCTCAAGAGGCTCTCTGGGTGGTCCATTTGTTTG	3853
Db	3782	CTTGCTTTCCCCCAGGCCATCTCTTACCTGAGGGCTCTCTGGGTGGACCATGTGTTTG	3841
Qy	3854	CCCTTCGGGGAACGTCTGTGGCGCTTTTCGGGCTGTGTGTGCAACCGGGGGTTCGCGAA	3913
Db	3842	CCCTTCGGGGAACGTCTGTAGGCACTTTTCGGGCTGTGTGTGCAACCGGGGGTTCGCGAA	3901
Qy	3914	GGGGTGGACTTCATACCGGTGAGTCTATGGAACAATACCAATGGGTCTCGGCTCTTCCAC	3973
Db	3902	GGGGTGGACTTCATACCGGTGAGTCTATGGAACAATACCAATGGGTCTCGGCTCTTCCAC	3961
Qy	3974	AGACAACCTCAACCCCCCGGCTGTACCGCAGACATTCAGAGTGGCACAATCTGCACGCTCC	4033

3962 AGACAACTCATCCCTCCGGCCGTACCGCAAAATTCAGAGTGGCACATTTACACGGCTCC 4021

4034 TACTGGCAGCGCAAGAGACACAAAGTCGCGCTCGGTATGCAGCCAAAGGTACAAGGT 4093

4022 CACTGGCAGCGCAGAGACACAAAGTCGCGCTGCATATGCAGCCCAAGGTACAGGT 4081

4094 GCTCGTCCGTGAAACCCGTCCTGTCGCCCACTTACGGTTTGGGGGWPATATGTCGAAGC 4153

4082 GCTCGTCCTAACCCGTCCTGTCGCCCACTTGGGCTTTGAGAGCTATATGTCGAAGC 4141

4154 ACACGATATGACCCATAACATCAGAACTCAGGAGTAAGACCAATACCAACGGCGGCTCCAT 4213

4142 ACATGGCATCGAGCCTTACATCAGAACTGGGTAGAACCATACCAACGGCGGCCCCAT 4201

4214 TACGTACTCCACCTPATGGCAAGTTCCTTCCCGAACGGTGGCTGTTCTGGGGGGCGCTATGA 4273

4202 CACGTACTCCACCTATTTGCAAGTTCCCTGGCAACGGTGGATGCTCCGGGGCGCTATGA 4261

4274 CATCATATATGTGATCAGTGCACATCACTCACTTCGACTACCAATCTTTGGGCAATCGGCAC 4333

4262 CATCATATATGTGATGAATGCCACTCACTCACTTCGACTCACTACATCTTTGGGCACATCGGCAC 4321

4334 AGTCCGTGACCAACGCGAGAGCGCTGGAGCGGGCTGCTGCTGTGTCGCCACACGCTACACC 4393

4322 AGTCCGTGATCAGCAGAGAGCGCTGGAGCGGGCTGCTGCTGTGTCGCCACACGCGCC 4381

4394 TCCGGATCGGTTACGCTGCCACACCCCAATATCAGAGAAATAGGCTGTTCACAACTGG 4453

4382 TCCGGGATCGATACCGCTGCCACACCCCAATCAGAGAAATGGCCCTGTCCACACTGG 4441

4454 AGAGATCCCTCTTATATGGCAAAAGCCATCCCATTTGAGGCCATCAAGGGGGGGAGGCATCT 4513

4442 AGAGATCCCTCTTATATGGCAAGCCATCCCATTTGAGGCCATCAAGGGGGGAAGGCATCT 4501

4514 CATTTTGTGCATTTCCAAAGAAATATGTGACGAGCTGCGCGCAAGCTGACAGGCTTCGG 4573

4502 CATCTTCTGCCATTTCCAAAGAAAGTGTGACAGCTGCGCGCAAGCTGACAGGCTTCGG 4561

4574 ACTGAAGCTGTAGCATATTTACCGGGGCTTCAGTGTGCGGTATACCGCTATCGGAGA 4633

4562 ACTCAATGCTGTAGCTATTTACCGGGGCTTCGATGTGTCCGTCAVACGCTAGCGGAGA 4621

4634 CTTGCTGTTCGTGCCAACAGACCTCTAATGACGGTTTCACCGGCGATTTGACTCAGT 4693

4622 CTTGCTGTTCGTGCCAACAGACCTCTAATGACGGTTTTCACCGGCGATTTGACTCAGT 4681

4694 GATGACTGCAATATCATGTGTCCACGACAGTGCATTTGAGCTTGAGTATCCACCTTCAC 4753

4682 GATGACTGCAACATATGTGTCCACGACAGTGCATTTGAGCTTGAGTATCCACCTTCAC 4741

4754 CATTCAGACGACGACCGTGCCCAAGACCGGCTGCGGCTCGCAACGGCGAGTAGAAC 4813

4742 CATTCAGACGACAAAGCTGCCCCAAGACCGGCTGCGGTGCGAGCGGGAGGTAGGAC 4801

4814 TGGCAGGGGTAGAGTGGCATCTAAGGTTTGTGACTCCAGAGAAAGCGCCCTCGGGCAT 4873

4802 TGGCAGGGGTAGAGTGGCATCTACAGGTTTGTGACTCCAGAGAAAGCGCCCTCAGGCAT 4861

4874 GTTTCGATCTTCGGTCTGTGTGAGTGTATCAACCGGGCTGCTGTTGATCAGGTTCAC 4933

4862 GTTTCGATCTTCGGTCTGTGTGAGTGTATGACGAGGCTGCGCTTGGTATCAGGTTCAC 4921

4934 GCCCGCTGAGACTCGGTTAGTGTGCGGCTTAACTAAATACACGAGGTTGCCGCTG 4993

4922 GCCCGCTGAGACTCGGTTAGTGTGCGGCTTAACTAAATACACGAGGTTGCCGCTG 4981

4994 CAGGACCACTCGGTTCTGGAGAGAGCTCTTACAGGCTCAGCCACATAGTCCCCA 5053

4982 CAGGACCACTAGTCTCGGAGAGAGCTCTTACAGGCTCAGCCACATAGTCCCCA 5041

5054 CTTCTGTCCCAAGCTAAACAGGACGAGACAATCTTCTTACCTGGTGGCATATCAAGC 5113

Dd 5042 CTTCTTGTCAGACCAAAACAGGACAGGACAAACCTCCCTACTCGTGTAGCATACCAAGC 5101
Qy 5114 TACAGTGTGCGCCAGGGCTCAAGCTCCACCTCCACCTGCTGGGACCAAAATGTAAGTGTCT 5173
Dd 5102 CACAGTGTGCGCCAGGGCTCAGGCTCCACCTCCATGCTGGGACCAAAATGTAAGTGTCT 5161
Qy 5174 CATACGGGTGAACACTTACATGCAACGGGCCAACACCCCTGCTGTATAGGCTTAGGACCGGT 5233
Dd 5162 CATACGGGTAAAGCCACACTGCATGGGCCAACGCCCTGCTGTACAGGCTAGGACCGGT 5221
Qy 5234 CCAAAATGAGGTCATCTTACACACACCCCACTAACTAAATACATCATGCGATGCAATGTCGGC 5293
Dd 5222 TCAAAATGAGGTCATCTTACACACACCCCACTAACTAAATACATCATGCGATGCAATGTCGGC 5281
Qy 5284 TGACCTGAGGTCGTCACTAGCACCTGGGTGTGCTAGGCGAGTCCCTTGCAGCTTTGGC 5353
Dd 5282 TGACCTGAGGTCGTCACTAGCACCTGGGTGTGCTAGGCGAGTCCCTTGCAGCTTTGGC 5341
Qy 5354 CGCATACTGCTGACGACAGGCAAGTGTGCTCATTTGTGGCAGGATCATCTTGTCCGGAA 5413
Dd 5342 CCGCTACTGCTGACGACAGGCAAGTGTGCTCATTTGTGGCAGGATCATCTTGTCCGGAA 5401
Qy 5414 GCCAGCTGCTGCTCCGACAGGAGTCTCTTACCAGAGTTCGATGAGATGGAAGTGTG 5473
Dd 5402 GCCAGCTGCTTATCCCGACAGGAGTCTCTTACCAGAGTTCGATGAGATGGAAGTGTG 5461
Qy 5474 TGCCTCACAACTTCCTTACATCGACAGGGAATGCAGCTCGCGGACCAATTCAGACAAA 5533
Dd 5462 TGCTTACACACTCCCTTACATCGACAGGAGTGTGCTCATTTGTGGCAGGATCATCTTGTCCGGAG 5521
Qy 5534 GCGCTCGGTTGTGCAAAACGGCACACAGCAAGCGGAGGTGTGCTGCTCCGTTGGTGA 5593
Dd 5522 GCGCTCGGATTTGTGCAAAACAGCCACCAAGCAAGCGGAGGTGTGCTGCTCCGTTGGTGA 5581
Qy 5594 GTCCAAAGTGGGAGCCCTTGTGAGACCTTGTGGCGAGACACATGTGGAATTCATCAGCGG 5653
Dd 5582 GTCCAAAGTGGGAGCCCTTGTGAGACCTTGTGGCGAGAACACATGTGGAATTCATCAGCGG 5641
Qy 5654 AATACAGTACCTAGCAGGCTTATCCACTCTGCTGGAAACCCCGATAGCATCATTTGAT 5713
Dd 5642 GATACAGTACTTGGCAGGCTTATCCACTCTGCTGGAAACCCCGATAGCATCATTTGAT 5701
Qy 5714 GGCATTTACAGTCTTATCATGAGCCCGGTACCAACCAAAACACCCCTGCTTTAAACAT 5773
Dd 5702 GGCATTTACAGCTCTATCACAGCCCGTACCAACCAAAATACCCCTGCTTTAAACAT 5761
Qy 5774 CTTGGGGGATGGGTGGTGGCCCACTCGCTCCCTCCAGCGGTGCTGAGCTTTGCTGGG 5833
Dd 5762 CTTGGGGGATGGGTGGTGGCCCACTCGCTCCCTCCAGCGGTGCTGAGCTTTGCTGGG 5821
Qy 5834 CGCCGATCGCCGAGCGGCTGTTGGCAGCATAGGCTTGGGAAGGTGCTCGTGGACAT 5893
Dd 5822 CGCCGATTTGCCGTGGCCGCTTGGCAGCATAGGTCTCGGAAGGTACTTGTGCAAT 5881
Qy 5894 CTTGGCGGCTATGGGACAGGGGTAGCCGGCACTCGTGGCTTTAAGGTATAGCGG 5953
Dd 5882 TCTGGCGGCTATGGGCGGGGTGGCTGGGCACTCGTGGCTTTAAGGTATAGCGG 5941
Qy 5954 CGAGTGGCCCTCCACCGAGGACCTGCTCACTTACTTCCCTGCCATCTCTCTCTGGTGC 6013
Dd 5942 CGAGATGCCCTCCACTGAGATCTGTTAATTTACTCCCTGCTTCTCTCTGGGCT 6001
Qy 6014 CTTGCTGCTGGGTCGTGTGCGACCAATATGCTGCGGACGTCGCGGCGGAGAGGG 6073
Dd 6002 CTTGCTGCTGGGTCGTGTGCGACCAATATGCTGCGGACGTCGCGGCGGAGAGGG 6061
Qy 6074 GGCTGTGAGTGGATGAAACCGGCTATAGGCTTGCCTTGGGGGTATACAGCTCCCC 6133
Dd 6062 GGCTGTGAGTGGATGAAACCGGCTATAGGCTTGCCTTGGGGGTATACAGCTCCCC 6121
Qy 6134 TACGACTATGTGCTGAGAGCGGCTGCAACAGTGTCTACTCAGATCCCTCTCTAGCTT 6193
Dd 6122 CACGACTATGTGCTGAGAGCGGCTGCAACAGTGTCTACTCAGATCCCTCTCTAGCTT 6181

Qy 6194 TACCATCACTCAACTGCTGAGCGGCTCCACAGTGGATTAATGAGGACTGCTTACGCC 6253
Dd 6182 TACCATCACTCACTGCTGAGAGAGGCTTCATCAGTGGATTAATGAGGACTGCTTACGCC 6241
Qy 6254 ATGCTCCGGCTCGTGGCTAAAGGATGTTTGGATATGATATGCAAGTGTGTTGACTGACTT 6313
Dd 6242 TTGTTCCGGCTCGTGGCTAAAGGATGTTTGGATATGATATGCAAGTGTGTTGAGTGAATT 6301
Qy 6314 CAGACCTGGCTCCAGTCCAAATCTCTGCGCGGTTACCGGAGTCCCTTTCTGTCTGATG 6373
Dd 6302 CAAGACTGGCTCCAGTCCAAAGTCTCTGCGCGGTTACCGGAGTCCCTTTCTGTCTGATG 6361
Qy 6374 CCAACCGGGTACAAAGGAGTCTGCGGGGGGACGGGATCATGCAAAACCACTCCGCCATG 6433
Dd 6362 CCAACCGGGTACAAAGGAGTCTGCGGGGGGATGCGATCATGCAAAACCACTCCGCCATG 6421
Qy 6434 CGGAGCACAGATCGCCGACATGTCAAAACCGTTCCATGAGGATCGTAGGGCTTAGAAC 6493
Dd 6422 TGGAGCACAGATCACCGGACATGTCAAAATGGTCCATGAGGATTTGTTGGGCCCAAAAC 6481
Qy 6494 CTGACGCAACAGCTGGGACGGAAAGTTCCTCCCATCAACGATACACCAAGGACCTTGAC 6553
Dd 6482 CTGACGCAACAGCTGGGACGGAAAGTTCCTCCCATCAACGATACACCAAGGACCTTGAC 6541
Qy 6554 ACCCTCCCGCGGCCCAACTATTCCAGGGGCTATGCGGGTGTGCTGAGGAGTACGT 6613
Dd 6542 GCCCTCCCGCGGCCCAACTATTCCAGGGGCTATGCGGGTGTGCTGAGGAGTACGT 6601
Qy 6614 GGAGTTACCGGTGTGGGGATTTCCATAGTACCGGGCATGACCACTGACAACTGATAA 6673
Dd 6602 GGAGTTACCGGTGTGGGGATTTCCATAGTACCGGGCATGACCACTGACAACTGATAA 6661
Qy 6674 GTGCCCATGCCAGGTTGCGGCCCGGAAATTTCTTCAAGGAGTGTGAGGAGTGGCTGCA 6733
Dd 6662 ATGCCCATGCCAGGTTGCGGCCCGGAAATTTCTTCAAGGAGTGTGAGGAGTGGCTGCA 6721
Qy 6734 CAGTACGCTCCCGCGGTGCAAACTCTTACGGGAGGACCTCACCTTCCAGGTCGGGCT 6793
Dd 6722 CAGTATGCTCCAGTGTGCAAACTCTTCTTACGAGAGGAGTGTGATTCAGGTCGGGCT 6781
Qy 6794 CAAACCAATACTGTGTCGGTGGAGCTCCCATGCGAGCCCGGAAACCGGACGTAACAGTGT 6853
Dd 6782 CAACCAATACTGTGTCGGTGGAGCTCCCATGAGCCCGGAAACCGGAGTGTGAGTGTGT 6841
Qy 6854 TACTTCCATGCTCACCGATCCCTCCACATTTACAGGAGAGACCGCTTAAGCTGAGCTGGC 6913
Dd 6842 CACTTCCATGCTCACCGACCCCTCTCATATTTACAGCAGAGACCGCTTAAGCTGAGCTGGC 6901
Qy 6914 TAGAGGCTTCCCGCTCTTTAGCCAGTCTCATCAGTGTAGCAGTGTGCTGCGCTTCTTT 6973
Dd 6902 CAGGGGCTTCCCGCTCTTTAGCCAGTCTTTAGCCAGTCTTTAGCTTGTGCTGCGCTTCTTT 6961
Qy 6974 GAAGGAGATGCACTACCAACCATGACTCCCGGACGCTGAGCTCATCTCAGGAGCAACCT 7033
Dd 6962 GAAGGAGATGCACTACCAACCATGACTCCCGGACGCTGAGCTCATCTCAGGAGCAACCT 7021
Qy 7034 CTTGTCGGGAGGAGTGGCGGAAACATCACTTCGCTGGAGTTCAGAGATTAAGGTAGT 7093
Dd 7022 CTTGTCGGGAGGAGTGGCGGAAACATCACTTCGCTGGAGTTCAGAGATTAAGGTAGT 7081
Qy 7094 AATTTCTGATCTTTTTCGAAACCGCTTTCAGCGGAGGGGATGAGAGGAGATATCCGTCGC 7153
Dd 7082 AATTTCTGATCTTTTTCGATCCGATTCGAGTTCGGGCGGTGGAGATGAGAGGAGATATCCGTCGC 7141
Qy 7154 GCGGAGATCTCTCGGAAATTCAGGAGTTCCTCCCTCAGCGTTCGCCATATGCGGACGCC 7213
Dd 7142 GCGGAGATCTCTCGGAAATTCAGGAGTTCCTCCCTCAGCGTTCGCCATATGCGGACGCC 7201
Qy 7214 GGACTACATCTCTCCACTGCTAGAGTCTCTGGAGGACCCGCACTTACCTCCCTCGGTGGT 7273
Dd 7202 GGACTACATCTCTCCACTGCTAGAGTCTCTGGAGGACCCGCACTTACCTCCCTCGGTGGT 7261

QY	7274	ACACGGATGCCCATTTGCCCACTTACCAAGGCTCTCCAAATACCACTCCACGGAGAAAGAG	7333
DB	7262	ACACGGTTCCTCTGACAGAAATCCAAATGCTCTTTGCTTTGCGGAGCTTCGGCATTAGAC	7333
QY	7334	GACGGTTCCTCTGACAGAAATCCAAATGCTCTTTGCTTTGCGGAGCTTCGGCATTAGAC	7333
DB	7322	GACGGTTCCTCTGACAGAGTCCACGGTCTCTTCTGCTTTGCGGAGCTTCGCTACTAAGAC	7381
QY	7394	CTTCGGTAGCTCCGGATCGTCCGGCCGTTGATAGGGCACGGCGACCGCCCTTCCTGACCT	7453
DB	7382	CTTTGGCAGCTCCGGTCTCGGCCGTGACAGCGGCACGCGGACTGGCCCTCCCGATCA	7441
QY	7454	GGCTCCGACGACGGTGAACAAGATCCGAGTTCGATCTTCCTTCCATGCCCCCT	7512
DB	7442	GGCTCCGACGACGGCGACAAGAGATCCGAGTTCGATCTTCCTTCCATGCCCCCT	7501
QY	7514	TGAGGGGACGGGGGACCCGATCTCAGCGACGGGTCTTGCTTACCGTTCGATGAGGA	7573
DB	7502	CGAGGGAGACCGAGGGACCCGACCTCAGCGACGGGTCTTGCTTACCGTTCGATGAGGA	7561
QY	7574	GGCTAGTAGGATTCGCTCTGCTCTGCTCAATGCTCTATACGTTCGACAGCGGCCCTGATCAC	7633
DB	7562	AGCTGGTTCGAGACGCTGCTCTGCTCTGCTCAATGCTCTATACGTTCGACAGCGGCCCTGATCAC	7621
QY	7634	GCCATGCGCTCGGAGGAAATGATGCTGCCCATCAACCGCTTGAGCAACTCTTTGCTGCG	7693
DB	7622	GCCATGCGCTCGGAGGAGACAAGTTGCCCATCAATCCGTTGAGCAACTCTTTGCTGCG	7681
QY	7694	TCACCAACAATGCTTCACGCCACAACAATCCGACGCGACGCTTCGGCAGAGAAGGT	7753
DB	7682	TCACCAACAATGCTTCACGCCACAACAATCCGACGCGACGCTTCGGCAGAGAAGGT	7741
QY	7754	CACCTTTGACAGATTCGAAGTTCCTGGATGATCAATTCGGGACGCTATCAAGGAGATGAA	7813
DB	7742	CACCTTTGACAGATTCGACGCTTCCTGGACCACTACCGGACGCTCTCAAGGAGATGAA	7801
QY	7814	GGCGAAGCGTCCACAGTTAAGCTTAAGCTTCTATATAGAGAGCGCTGCAAGCTGAC	7873
DB	7802	GGCGAAGCGTCCACAGTTAAGCTTAAGCTTCTATATAGAGAGCGCTGCAAGCTGAC	7861
QY	7874	GCCCCACATTCGSCCAAAATCCAAATTTGGCTATGCGGCAAGGACGTCGGAACCTATC	7933
DB	7862	GCCCCACATTCGSCCAAAATCCAAATTTGGCTATGCGGCAAGGACGTCGGAACCTATC	7921
QY	7934	CAGCAGGCGCTTAAACCACTCCGCTCCGCTGCTGGGAGGACTTGTCTGGAAGACACTGAAAC	7993
DB	7922	CAGCAGGCGCTTAAACCACTCCGCTCCGCTGCTGGGAGGACTTGTCTGGAAGACACTGAAAC	7981
QY	7994	ACCAATTGACACACCACTCATGGCAAAAAGTAGGTTTTCTGGTCCAAACGAGAGGG	8053
DB	7982	ACCAATTGACACACCACTCATGGCAAAAAGTAGGTTTTCTGGTCCAAACGAGAGAGG	8041
QY	8054	AGSCCGCAAGCAGCTCGCTTATGCTATTCGACAGCTCCGAGTTCTGCTATGCGAGAA	8113
DB	8042	AGSCCGCAAGCAGCTCGCTTATGCTATTCGACAGCTCCGAGTTCTGCTATGCGAGAA	8101
QY	8114	GATGGCCCTTTACGACGTGTCTCCACCCCTTCCTCAGGCCGTCATGGGCTCTCATACGG	8173
DB	8102	GATGGCCCTTTACGACGTGTCTCCACCCCTTCCTCAGGCCGTCATGGGCTCTCATACGG	8161
QY	8174	ATTTCATATCTCCCAAGCAGCGGTTCGAGTTCTCTGCTGAAATACCTGGAAATCAAGAA	8233
DB	8162	ATTTCATATCTCTCTGGGAGCGGTTCGAGTTCTCTGCTGAAATACCTGGAAATCAAGAA	8221
QY	8234	ATGCCCTATGGCTTCCTCATATGACACCCGCTGTTTGACTCAACCGTCACTGAGATGA	8293
DB	8222	ATGCCCTATGGCTTCCTCATATGACACCCGCTGTTTGACTCAACCGTCACTGAGATGA	8281
QY	8294	CATTCTGCTTTGAGAGTCAATTTACCAATGTTGTGACTTGGCCCCCGAGGCCAGACGGC	8353
DB	8282	CATCCGTACTGAGGAATCAATTTACCAATGTTGTGACTTGGCCCCCGAGGCCAGACGGC	8341
QY	8354	CATTAAGTCTCTCAGAGCGGCTTTACATCGGGGGTCCCTGACTAACCTAAAGGGCA	8413

Db	8342	CATAAGGTTCGCTCAACAGAGCGGCTTTATGTCGGGGTCCCTGACTAATTCGAAGGGCA	8401
Qy	8414	GAACTGCGGTATCGCGGTTCGCCGAGAGTGGCGTGCTGACGACTAGCTGCGGTAAATAC	8473
Db	8402	GAACTGCGGTATCGCGGTTCGCCGAGAGTGGCGTGCTGACGACTAGCTGCGGCAAC	8461
Qy	8474	CCTCACATGTTACTTGAAGGCCACTGCAGAGCTGTCGAGCTGCAAAAGCTCCAGACTGCAC	8533
Db	8462	CCTCACATGTTACTTGAAGGCCACTGCAGAGCTGTCGAGCTGCAAAAGCTCCAGACTGCAC	8521
Qy	8534	GATGCTGTTGAACGGAGACGACCTTTGTCGTTATCTGTGAAAGCGCGGAAACCCAGAGGA	8593
Db	8522	GATGCTGTTGAACGGAGACGACCTTTGTCGTTATCTGTGAGAGTGGGGAACCCAGAGGA	8581
Qy	8594	TGCGGCGCCCTACGAGCCTTCAGGAGGCTATGACTAGGTATTCGCCCCCCCCGGGA	8653
Db	8582	TGCGGCGCCCTACGAGCCTTCAGGAGGCTATGACTAGGTATTCGCCCCCCCCGGGA	8641
Qy	8654	TCCGCCCAACAGAAATACGACCTGGAGCTGATAACATCATGTTCTCTCCAAATGTCAGT	8713
Db	8642	CCCGCCCCAACAGAAATAGACCTTGGAGCTGATAACATCATGTTCTCTCCAAATGTCAGT	8701
Qy	8714	CGCGCAGATGATCTCGGCAAAAGGATATCTACTCTCAACCGTGAACCCACACCCCCCT	8773
Db	8702	CGCGCAGATGATCTCGGCAAAAGGATATCTACTCTCAACCGTGAACCCACACCCCCCT	8761
Qy	8774	TGCACGGGCTGCGTGGAGACAGTACACACTCCAAATCAACTCTTGCTAGGCAATAT	8833
Db	8762	CGCACGGGCTGCGTGGAGACAGTATAGACACTCCAGTCACTCTTGCTAGGCAATAT	8821
Qy	8834	CATCATGATGCGGCCACCCCTATGGGCAAGATGATTTCTGATGACTCACTTTTCTCCAT	8893
Db	8822	CATCATGATGCGGCCACCCCTATGGGCGAGATGATTTCTGATGACTCACTTTCTCTCAT	8881
Qy	8894	CCTTCTAGCTCAAGAGCAACTTGAAAGCCCTGGATTTGTCAGATCTACGGGCTTGCTA	8953
Db	8882	CCTTCTAGCTCAGAGCAACTTGAAAGCCCTGGATTTGTCAGATCTACGGGCGCTTGTTA	8941
Qy	8954	CTCCATTGAGCCACTGACCTCAGATCATTTGAACGACTCCATGGTCTTAGCGCATT	9013
Db	8942	CTCCATTGAGCCACTGACCTACTCAGATCATTTGAACGACTCCATGGTCTTAGCGCATT	9001
Qy	9014	TACACTCAAGTTPACTCTCCAGTGAGATCAATPAGGTGGTCTCATGCTCAGGAACT	9073
Db	9002	TTCACTCCACAGTTACTCTCCAGTGAGATCAATPAGGTGGTCTCATGCTCAGGAACT	9061
Qy	9074	TGGGTTACCAACCTTGGACCTTGAGACATCGGGCCAGAGTGTCCGCTTAGCTACT	9133
Db	9062	TGGGTTACCCCTTTTGGAGCTTGAGACATCGGGCCAGAGTGTCCGCTTAGCTACT	9121
Qy	9134	GTCCAGGGGGGAGGGCCGCCCTTTGTGGCAGATACCTCTTTAACTGGGCGATGAGGAC	9193
Db	9122	GTCCAGGGGGGAGGGCTGCCACTTGCGCAAGTACTCTTCACTGGGCGATGAGGA	9181
Qy	9194	CAAGCTTAAACTCATCTCAATCCCGCGCGGTCCACAGCTGGAGCTGTGTGGCTGGTTGCT	9253
Db	9182	CAAGCTTAAACTCATCTCAATCCCGCGGTCCACAGCTGGAGCTGTGTGGCTGGTTGCT	9241
Qy	9254	CGCTGGTTACAGCGGGGAGACATATACAGCCTGTCTGTCGCCGACCCCGCTGGTT	9313
Db	9242	TGCTGGTTCAAAGGGGAGACATATACAGCCTGTCTGTCGCCGACCCCGCTGGTT	9301
Qy	9314	TCCGTTGTGCTACTCTACTTTCTCTAGGGGTAGGCATTTACCTGCTCCCCAACCGATG	9373
Db	9302	CATGTTGTGCTACTCTACTTTCTCTAGGGGTAGGCATCTACCTGCTCCCCAACCGGTG	9361
Qy	9374	AAGGGAGCTAACCACTCCAGGCTTAAGCCATTTCTGTTTTTTTTTTT	9424
Db	9362	AAGGGAGCTAACCACTCCAGGCAATAGGCCATTTCCCTTTTTTTTTTTT	9412

AAQ81559	AAQ81559 standard; DNA; 9413 BP.
ID	AAQ81559 standard; DNA; 9413 BP.
XX	
XX	
AC	AAQ81559;
XX	
DT	06-DEC-1995 (first entry)
XX	
DE	Hepatitis C virus helicase gene.
XX	
KW	Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase;
KW	baculovirus; recombinant production; ss.
XX	
OS	Hepatitis C virus.
XX	
Key	Location/Qualifiers
FH	330..9362
FT	/*tag= a
FT	9382..9387
FT	/*tag= C
FT	/rpt_type= inverted
FT	9390..9395
FT	/*tag= d
FT	/rpt_type= inverted
XX	
PN	JPC6319583-A.
XX	
PD	22-NOV-1994.
XX	
PF	18-SEP-1992; 92JP-00249241.
XX	
PR	18-SEP-1992; 92JP-00249241.
XX	
PA	(SOYA-) SOYAKU GLUTSU KENKYUSHO KK.
XX	
DR	WPI; 1995-040330/06.
DR	P-PSDB; AAR68864.
XX	
PT	of hepatitis C virus helicase gene in baculovirus - useful for large
PT	scale prodn. of RNA helicase.
XX	
PS	Claim 1; Fig 1-4; 9pp; Japanese.
XX	
CC	AAQ81559 encodes AAR68864 hepatitis C virus (HCV) RNA helicase. The DNA
CC	was used in the construction of an expression vector, which was used to
CC	transform a baculovirus host. The transformed baculovirus could then be
CC	used for the recombinant prodn. of HCV RNA helicase
XX	
SQ	Sequence 9413 BP; 1886 A; 2817 C; 2693 G; 2017 T; 0 U; 0 Other;
Query Match	
Best Local Similarity 85.6%; Score 8307; DB 2; Length 9413;	
Matches 872; Conservative 0; Mismatches 690; Indels 0; Gaps 0	
Qy	14 TGGGGGGGACACTCCACCATGAATCAGTCCCTGTGAGGAATCTGCTTTCACGCAGAA 73
Db	2 TGGGGGGGACACTCCACCATGATCAGTCCCTGTGAGGAATCTGCTTTCACGCAGAA 61
Qy	74 AGCGTCTAGCATGGCGTTAGTATGAGTGTGTGTGAGCGCTTCAGAACCCCGCTCCCGGG 133
Db	62 AGCGTCTAGCATGGCGTTAGTATGAGTGTGTGTGAGCGCTTCAGAACCCCGCTCCCGGG 121
Qy	134 AGAGCCATAGTGGTCTCGGGAACCGGTGAGTACACCGGAATTGCCAGACGACCGGGTCC 193
Db	122 AGAGCCATAGTGGTCTCGGGAACCGGTGAGTACACCGGAATTGCCAGACGACCGGGTCC 181
Qy	194 TTTCTTTGATCAACCGCTCAATGCTTCGAGATTGGCGCTGCCCGCGAGACTGCTAG 253
Db	182 TTTCTTTGATCAACCGCTCAATGCTTCGAGATTGGCGCTGCCCGCGAGACTGCTAG 241
Qy	254 CCGAGTACTGTGGGTCCGGAAGCCCTGTGGTACTGCTGATAGGTGCTTGGCGAGTG 313
Db	242 CCGAGTACTGTGGGTCCGGAAGCCCTGTGGTACTGCTGATAGGTGCTTGGCGAGTG 301

Qy	314	CCCCGGAGGTCGTGAGACCGCGTGCACCATGAGAGCAAGATCCTTAAACCTCAAGAAAAAC	373
Db	302	CCCCGGAGGTCGTGAGACCGTGCATCATGAGCAAAATCCTTAAACCTCAAGAAAAAC	361
Qy	374	CAAAAGTAAACACCAACCGCCGCCACAGGACGTCAGTTCCCGGGCGGTGGTCAAGTCGT	433
Db	362	CAAAAGTAAACCAACCGCCGCCACAGGACGTTTAAAGTCCCGGGCGGTGGTCAAGTCGT	421
Qy	434	TGCTGGAGTTTAACTGTTGCGGCAGGGGCCCCAGGTTGGGTGTGTCGCGCACTAGGAA	493
Db	422	TGCTGGAGTTTAACTGTTGCGGCAGGGGCCCCAGGTTGGGTGTGTCGCGCACTAGGAA	481
Qy	494	GGCTTCGAGCGGTGCGAACTCGTGGAAAGGCGAACACTTATCCAAAGGCTCGCCGACC	553
Db	482	GACTTCGAGCGGTGCGAACTCGTGGAAAGGCGAACACTTATCCCAAGGCTCGCCGACC	541
Qy	554	CGAGGCGAGGCTTGGCTCAGCCCGGGTACCTTGGCCCTCTATGGCAATGAGGCGCT	613
Db	542	CGAGGCTAGGACTTGGCTCAGCCCGGGTACCTTGGCCCTCTATGGCAATGAGGCGTAT	601
Qy	614	GGGCTGGGAGGATGGCTCCTGTACACCCCGGCTCCCGGCTAGTTGGGGCCCCCAGGA	673
Db	602	GGGCTGGGAGGATGGCTCCTGTACACCCCGTGGCTCTCGGCTAGTTGGGGCCCCCAGGA	661
Qy	674	CCCCCGGCTAGTFCGCGTAACTTGGGTAAAGTTCATCGATACCCCTTACATGCGGCTTCGC	733
Db	662	CCCCCGGCTAGTFCGCGTAACTTGGGTAAAGTTCATCGATACCCCTTACATGCGGCTTCGC	721
Qy	734	CGATCTCATGGGTACATTCGCTCGTGGCGGCCCGCTAGGGGGCGCTGCGCAGGCGCTT	793
Db	722	CGACCTCATGGGTACATTCGCTCGTGGCGGCCCGCTAGGGGGCGCTGCGCAGGCGCTT	781
Qy	794	GGCACCGGTGTCGGGTTCTGAGAGACGCGTGCACTATGCAACAGGGAACTTGCGCCGG	853
Db	782	GGCACATGGTGTCCGGTTCTGAGAGACGCGTGCACTATGCAACAGGGAATCTGCCCGG	841
Qy	854	TTGCTCTTTCTCTATCTTCTCTTGGCTGTGCTGTCTGTGTTGACCATCCAGCTTCCGC	913
Db	842	TTGCTCTTTCTCTATCTTCTCTTGGCTGTGCTGTGTTGACCATCCAGCTTCCGC	901
Qy	914	TTATGTAAGTCGCAACGTGTCCGGATATACATGTCAAGAAAGACTGTCCCACTCAAG	973
Db	902	TTACGAGGTGCGCAACGTGTCCGGATATACCATGTCAAGAAAGACTGTCCCACTCAAG	961
Qy	974	CATTGTGATGAGCAGCGGACGTGATCATGCACTATCCCGGCTGGGTGCCCTGTGTCA	1033
Db	962	TATTGTGATGAGCAGCGGACATGATCAAGCAACCCCGGGTGGTGCCTCGGTCGG	1021
Qy	1034	GGAGGTTACAGCTCCCGTTGCTGGGTAGCGGTCACTCCACGCTCGCGGCCAGGAATGC	1093
Db	1022	GGAGAGTAATTTCTCCGTTGCTGGGTAGCGGTCACTCCACGCTCGCGGCCAGGAACAG	1081
Qy	1094	CAGGTCCTCCACTACGACAATAAGAGCCACGTCGACTGCTGTTGGGACGCTGCTTT	1153
Db	1082	CAGCATCCCAACACGACAATAGAGCCACGTCGATTTGCTCETTTGGGGCGCTGCTCT	1141
Qy	1154	CTGCTCGCTATGTAGCTGGGGATCTCTCGGATCTATTTCCTGGTCTCCACAGCTGTT	1213
Db	1142	CTGTTTCGCTATGTAAGTTGGGATCTCTGGGATCGGTTTTTCTGCTCTCCAGCTGTT	1201
Qy	1214	CACCTTCTCGCTCGCGGGAATGAGACATGTCAGGACTGCAATGTGCTCAATCTATCCCGG	1273
Db	1202	CACCTTCTCACCTCGCGGTAAGAGACGTTACAGATTTGCAATGTGCTCAATCTATCCCGG	1261
Qy	1274	CCATGTATCAGGTCACGGCATGCTTTGGGATATGATGTGAACTGGTTCACCTACACAGC	1333
Db	1262	CCACGTATCAGGTCACGGCATGCTTTGGGATATGATGTGAACTGGTTCACCTACACAGC	1321
Qy	1334	CTTAGTGTGTGCAAGTTGCTCCGGATCCCAAGGCTCGTGTGCAATGTTGCGGGGGC	1393
Db	1322	CTTAGTGTGTATCGAGTACTCCGGATCCCAAGCGCTGCTGTGCAATGTTGCGGGGGC	1381
Qy	1394	CCACTGGGAGTCTTGGCGGGCTTGCCTACTATTCCATGCTAGGGAACCTGGGCTAAGT	1453

1382	Db	CCACTGGGGTGCCTTACGGGCGCTTGGCTACTATTCCATGGTGGGAACATGGGCTAAGGT	1441
1454	QY	TCTGATTGTGGCGCTACTCTTTCCGGCGGTTGACGGGGAGACCCACACGACGGGGAGGGT	1513
1442	Db	CTTGAATTGTGATGCTACTCTTTGCTGGCGTTTGACGGGCACACCCACGTCGACAGGGGAAG	1501
1514	QY	GGCGGGCCACACACCTTCGGGTTTCAGTCCCTTTTCTCTCTGGGCGGTCCTCAGAAAAAT	1573
1502	Db	GGTAGGCTCCAGCACCCAGAGCCTCGGTGTCTTGCTCTCTCAAGAGGCCATCTCTCAGAAAAAT	1561
1574	QY	CGAGCTTGTGAATACCAACGGCAGCTGGCAGACTCAACAGGAGCTTGCCCTAAATTGCAATGA	1633
1562	Db	CCAACCTGTGAACACCAACGGCAGCTGGCAGACTCAACAGAGACCGCTCTGAAATTGCCAATGA	1621
1634	QY	CTCCCTCCAAACTGGGTTCTTTGCGCGCTGTTTATCGCACACAGGTTCAACTCGCTCCGG	1693
1622	Db	CTCCCTCCAAACTGGGTTCTTTGCTGGCGCTGTTCTACGCAACAGGTTCAACGCGTCCGG	1681
1694	QY	GTGCGCGGACGCATGCGCAGCTGCCGCCCAATTGACTGGTTCCGCCAGGGGTGGGGCCC	1753
1682	Db	GTGCGCGGACGCATGCGCTAGCTGCCGCCCATCGATGAGTTCCGTCAGGGGTGGGGTCC	1741
1754	QY	CATCACCTATACTAAGCCTAACAGCTCGGATCAGAGGCGCTATTCTGGCGAATTACGGGCC	1813
1742	Db	CATCACCTATGATATGCTGAGAGCTCGGACCGAGGCCCATATTGCTGGCACTACGCGCC	1801
1814	QY	TCGACCGTGGTGTGCTACCGCGTCGCGAGGTGTGGTCCAGTGTATTGTTTACCCCC	1873
1802	Db	TCGACCGTGGGNAATCGTGCTGGTCGCGAGGTGTGGTCCAGTGTATTGTTTACTTCC	1861
1874	QY	AAGCCCTGTTGTGGTGGGGACCAACCGATCGTTCCGGTGTCCCTACGTTATAGCTGGGGGA	1933
1862	Db	GAGCCCTGTTGTAGTGGGGACGACCGATCGTTTCGGCGCTCCTACGTATTAGCTGGGGGA	1921
1934	QY	GAATGACACAGACGTGTATGCTCTCAACACACGCGTCCGCCACAGGCAACTGGTTTCGG	1993
1922	Db	GAATGACACAGACGTGTGCTGCTACTTAGCAACACGCGCCGCGCTCAAGGCAACTGGTTTGG	1981
1994	QY	CTGTACATGATGAATAGTACTGGGTTCACTAAGAAGTGGGAGGTCCCCCGTGTAAACAT	2053
1982	Db	GTGCACTGTGAATGAACAGCACTGGGTTCAACAGACGTGGGGGGCCCTCGGTGCAACAT	2041
2054	QY	CGGGGGGTCCGTTAACCCGACCTTGATCTGCCCCACGGAATGCTTCGGGAAGCACCCCGA	2113
2042	Db	CGGGGGGTCCGCCAACCAACACTTGGTCTGCCCCACGGAATGCTTCGGGAAGCACCCCGA	2101
2114	QY	GGCTACTTACACAAATGTGGCTCGGGGCCCTGGTTGACACCTAGTGTCCCTAGTAGCTA	2173
2102	Db	GGCCACTTACACAAAGTGTGGCTCGGGGCCCTGGTTGACACCCAGGTGCAATGTTGACTA	2161
2174	QY	CCCATACAGGCTTGGCACTACCCCTGCATCTCAATTTTCCATCTTTAAGGTTAGCAT	2233
2162	Db	CCCATACAGGCTTGGCACTACCCCTGCACGTGTTAACTTTACCGTCTTTAAGGTCAGGAT	2221
2234	QY	GTATGTGGGGCGGTGGAGCAACGGCTCAATGCCGATGCAATTAAGACTCGAGGAGACGG	2293
2222	Db	GTAATGTGGGGCGGTGGAGCACAGGCTCAATGTGCAATGCAATTTGGACTCGAGGAGACGG	2281
2294	QY	CTGTAACTTGGAGACACGGGATAGGTGAGAACTCAGCCCGCTGTGCTGTCTACACAGA	2353
2282	Db	CTGTGACTTGGAGACACGGGATAGGTGAGAACTCAGCCCGCTGTGCTGTCTACACAGA	2341
2354	QY	GTGGCAGATATGCCCCGTGTTTACCACCCCTACCGGCTTATACCACTGGTTGATCCCA	2413
2342	Db	GTGGCAGATATGCCCCGTGTTCTTCAACACCCCTACCGGCTTATACCACTGGCTTGAATCCA	2401
2414	QY	TCTCCATCAGAAACATCGGTGACGTGCATACTGCTACGGTGTAGGGGTACAGGCTTGTCTC	2473
2402	Db	TCTTACCGGAACATCGTGGACGTGCATACTGTACGGTATAGGTCGCACTGTGTCTC	2461
2474	QY	CTTTGCCAATCAAAATGGGAGTACATCCTGTTGCTTTCTCTCTCTGGCAGACGCGCGGT	2533

QY 3614 CGGTCCAAAGGTTCCATCACCACCAAAATGTACACCAATGTAGACCTTGACCTCGTGGCTG 3673
Db 3602 CGGCCCCGAAGGCTCCAAATCACCCCAAAATGTACACCAATGTAGACCAAGCTCGTGGCTG 3661
QY 3674 GCAGCGCCCCCGGGGGCGGCTCCATGACACCAATGACGTGTGGCAGCTCGGACCTTTA 3733
Db 3662 GCCGGCCCCCGGGGGCGGCTCCATGACACCAATGACGTGTGGCAGCTCGGACCTTTA 3721
QY 3734 CTTGGTCAAGAGACATGCTGATGATTCCTCGGTGCGCGGCGAGGCGACACAGAGGGAAG 3793
Db 3722 CTTGGTCAAGAGGATGCTGATGCTTCCTCGGTGCGCGGCGGCGACACAGAGGGAAG 3781
QY 3794 TCTACTCTCCCGCGGCGGCTCTCTACTCTGAAGGCTCTCGGCTGCTGCTGCTGCTGCTG 3853
Db 3782 CTTGCTTTTCCCGCGGCGGCTCTCTACTCTGAAGGCTCTCGGCTGCTGCTGCTGCTGCTG 3841
QY 3854 CTTTTCGGGGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3913
Db 3842 CTTTTCGGGGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3901
QY 3914 GCGGTGGAATTCATACCCGTTGATCTATGGAACCTACCAATGCGGTCTCCGGTCTTCC 3973
Db 3902 GCGGTGGAATTCATACCCGTTGATCTATGGAACCTACCAATGCGGTCTCCGGTCTTCC 3961
QY 3974 AGCAACTCAACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4033
Db 3962 AGCAACTCAACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4021
QY 4034 TACTGGGACGGGCAAGACACCAAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4093
Db 4022 CACTGGGACGGGCAAGACACCAAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4081
QY 4094 GCTGCTCTGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4153
Db 4082 GCTGCTCTGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4141
QY 4154 ACAGGTATCGACCTCAACCTGACAACTGGGTGAGGACCAATTACCAAGGCGGCTCCAT 4213
Db 4142 ACATGATCGAGGCTCAACCTGACAACTGGGTGAGGACCAATTACCAAGGCGGCTCCAT 4201
QY 4214 TACGTACTCCACTATGCAAGTTCCTTTCGCGACGGTGGCTGCTGCTGCTGCTGCTGCTG 4273
Db 4202 CAGGTACTCCACTATGCAAGTTCCTTTCGCGACGGTGGCTGCTGCTGCTGCTGCTGCTG 4261
QY 4274 CATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4333
Db 4262 CATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4321
QY 4334 AGTCTGGAACCAAGGAGAGGCTGAGCGGCTCGTGGCTCGTGGCTCGTGGCTCGTGGCTCG 4393
Db 4322 AGTCTGGAACCAAGGAGAGGCTGAGCGGCTCGTGGCTCGTGGCTCGTGGCTCGTGGCTCG 4381
QY 4394 TCCGGGATCGGTTACCGTGGCACACCCCAATATCGAGGAATAGGCTGCTGCTGCTGCTGCT 4453
Db 4382 TCCGGGATCGATCACGTTGGCACACCCCAATATCGAGGAATAGGCTGCTGCTGCTGCTGCT 4441
QY 4454 AGAGATCCCTTCTATGGAAGGCAATCCCATTTGAGGCTCAAGGGGGGAGGCTGCTGCTG 4513
Db 4442 AGAGATCCCTTCTATGGAAGGCAATCCCATTTGAGGCTCAAGGGGGGAGGCTGCTGCTG 4501
QY 4514 CATTTTCTGCAATCCAAAGAGATGTCAGAGCTCCGCGAAGGCTGACAGGCTCGG 4573
Db 4502 CATTTTCTGCAATCCAAAGAGATGTCAGAGCTCCGCGAAGGCTGACAGGCTCGG 4561
QY 4574 ACTGAACGCTGAGGATATATACCGGGGCTTGTATGCTGCTGCTGCTGCTGCTGCTGCTG 4633
Db 4562 ACTGAACGCTGAGGATATATACCGGGGCTTGTATGCTGCTGCTGCTGCTGCTGCTGCTG 4621
QY 4634 GTCGTTGCTGAGGCAACAGAGCTCTATGAGCGGCTTACCGGGGATTTTGAATCAGT 4693
Db 4622 GTCGTTGCTGAGGCAACAGAGCTCTATGAGCGGCTTACCGGGGATTTTGAATCAGT 4681

QY 4694 GATCGACTGCAATACATGT 4753
Db 4682 GATCGACTGCAATACATGT 4741
QY 4754 CATGTGACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4813
Db 4742 CATGTGACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4801
QY 4814 TGGCAGGGGTAGGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4873
Db 4802 TGGCAGGGGTAGGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4861
QY 4874 GTTCGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4933
Db 4862 GTTCGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4921
QY 4934 GCGCGTGAACCTCGGTTAGGTTGGGGCTTACCTTAAATACACGAGGTTGCCGCTGCTG 4993
Db 4922 GCGCGTGAACCTCGGTTAGGTTGGGGCTTACCTTAAATACACGAGGTTGCCGCTGCTG 4981
QY 4994 CCAGGACATCTGGAGTCTGGGAGAGGCTGCTTACAGAGGCTCACCACATAGATGCCA 5053
Db 4982 CCAGGACATCTGGAGTCTGGGAGAGGCTGCTTACAGAGGCTCACCACATAGATGCCA 5041
QY 5054 CTTCTCTCCAGACTAAACAGGAGGAGACACTTTTCTTACCTGCTGGCTATATCAAGC 5113
Db 5042 CTTCTCTCCAGACTAAACAGGAGGAGACACTTTTCTTACCTGCTGGCTATATCAAGC 5101
QY 5114 TACAGTGTGCGCAGGCTCAAGTCCACCTCATGCTGGGACCAATGTTGGAAGTGTCT 5173
Db 5102 CACAGTGTGCGCAGGCTCAAGTCCACCTCATGCTGGGACCAATGTTGGAAGTGTCT 5161
QY 5174 CATACGCTGAAACCTTACACTGACGCGGCAACACCTGCTGTATAGGCTAGGAGCGGT 5233
Db 5162 CATACGCTGAAACCTTACACTGACGCGGCAACACCTGCTGTATAGGCTAGGAGCGGT 5221
QY 5234 CCAAAATGAGGTCTATCTCTCACACCCCAATAAATTAATATCATATGCTGCTGCTGCTG 5293
Db 5222 TCANAATGAGGTCTATCTCTCACACCCCAATAAATTAATATCATATGCTGCTGCTGCTG 5281
QY 5294 TGACCTGAGGTGTGTCTAGACCTGGGTGCTGGGTGAGGAGGAGTCTTGTGAGCTTTGGC 5353
Db 5282 TGACCTGAGGTGTGTCTAGACCTGGGTGCTGGGTGAGGAGGAGTCTTGTGAGCTTTGGC 5341
QY 5354 CGCATCTGCTGACGACGAGGAGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5413
Db 5342 CGCATCTGCTGACGACGAGGAGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5401
QY 5414 GCGAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5473
Db 5402 GCGAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5461
QY 5474 TGCCTCAAACTTCTTACATGAGGAGGAGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 5533
Db 5462 TGCCTCAAACTTCTTACATGAGGAGGAGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 5521
QY 5534 GCGCTGCGGTTGTGTCAACCGGCAACAGAGGAGGAGGAGTGTGTCTGCTGCTGCTGCTGCTG 5593
Db 5522 GCGCTGCGGTTGTGTCAACCGGCAACAGAGGAGGAGGAGTGTGTCTGCTGCTGCTGCTGCTG 5581
QY 5594 GTCCAAAGTGGGAGGCTTGTGAGACTTCTTGTGGGCAAGCAATGTTGGAATTTTCAACAG 5653
Db 5582 GTCCAAAGTGGGAGGCTTGTGAGACTTCTTGTGGGCAAGCAATGTTGGAATTTTCAACAG 5641
QY 5654 AATACAGTACCTAGAGGCTTGTGAGACTTCTTGTGGGCAAGCAATGTTGGAATTTTCAACAG 5713
Db 5642 AATACAGTACCTAGAGGCTTGTGAGACTTCTTGTGGGCAAGCAATGTTGGAATTTTCAACAG 5701
QY 5714 GGCATTTTACGCTTATATCTAGCGGCTCACCACCAACCAACCTTCTGCTGCTGCTGCTGCTG 5773
Db 5702 GGCATTTTACGCTTATATCTAGCGGCTCACCACCAACCAACCTTCTGCTGCTGCTGCTGCTG 5761
QY 5774 CTTGGGGGAGTGGGTGGGCTGCCCAACTGCTCTCCAGCGCTGCTGCTGCTGCTGCTGCTG 5833

QY 7994 ACCAATTGACACCAATCATATGGCAAAAAAGTGAAGTTTCTGCGTCCAAACAGAGAAAGG 8053
 Db 7982 ACCAATTGATACCAATCATATGGCAAAAAAGTGAAGTTTCTGCGTCCAAACAGAGAAAGG 8041
 QY 8054 AGGCGCGAAGCAGATCGCTCCCTTATCGTATATCCAGACCTGGAGTTCGTATGCGAGAA 8113
 Db 8042 AGGCGCGAAGCAGATCGCTCCCTTATCGTATATCCAGACCTGGAGTTCGTATGCGAGAA 8101
 QY 8114 GATGCGCCCTTTACGACGTGGTCTCCACCTTCCTCAGGCCGTGATGGCTCCTCATACGG 8173
 Db 8102 GATGCGCCCTTTACGACGTGGTCTCCACCTTCCTCAGGCCGTGATGGCTCCTCATACGG 8161
 QY 8174 ATTTCAATACTCCGCCAAGCAGGAGTTCGTATATCCAGACCTGGAGTTCGTATGCGAGAA 8233
 Db 8162 ATTTCAATACTCCGCCAAGCAGGAGTTCGTATATCCAGACCTGGAGTTCGTATGCGAGAA 8221
 QY 8234 ATGCGCCCTATGCGCTTCTCATATGACACCCCGCTGTTTGACTCAAGGTCATGTAGATGA 8293
 Db 8222 ATGCGCCCTATGCGCTTCTCATATGACACCCCGCTGTTTGACTCAAGGTCATGTAGATGA 8281
 QY 8294 CATTCGTGTTGAGGAGTCAATTTTACCAATGTTGACTTTGSCCCCGCAGGCCAGACAGGC 8353
 Db 8282 CATTCGTGTTGAGGAGTCAATTTTACCAATGTTGACTTTGSCCCCGCAGGCCAGACAGGC 8341
 QY 8354 CATTAAGTTCGCTACAGAGCGGCTTTAATCGGGGTCCTGACTAACTCAAAAGGCA 8413
 Db 8342 CATTAAGTTCGCTACAGAGCGGCTTTAATCGGGGTCCTGACTAACTCAAAAGGCA 8401
 QY 8414 GAACCTCGGTTATCGCGGTCGCCGCGCAGTGGCGTGTGCTGACGACTAGCTGCGGTATATAC 8473
 Db 8402 GAACCTCGGTTATCGCGGTCGCCGCGCAGTGGCGTGTGCTGACGACTAGCTGCGGCAACAC 8461
 QY 8474 CCTCACATGTTACTTGAAGCCACTGACGCTGTCGAGCTGCAAAAGCTCCAGGACTCCAC 8533
 Db 8462 CCTCACATGTTACTTGAAGCCACTGCGGCTGTCGAGCTGCAAAAGCTCCAGGACTCCAC 8521
 QY 8534 GATGCTGTGAACGAGAGACCTTGTGTTATCTGTGAAGCGCGGAACCCAGGAGGA 8593
 Db 8522 GATGCTGTGAACGAGAGACCTTGTGTTATCTGTGAAGCGCGGAACCCAGGAGGA 8581
 QY 8594 TGGCGGCGCCTTAGAGCCCTTACGAGGCTATGACTAGTATTCGCCGCCGCCCGCGGGA 8653
 Db 8582 TGGCGGCGCCTTAGAGCCCTTACGAGGCTATGACTAGTATTCGCCGCCGCCCGCGGGA 8641
 QY 8654 TCGGCCCAACCAAGAAATACGACCTGAGCTGATAACATCATGTTCTCCCAATGTTCAGT 8713
 Db 8642 CCGCGCCCAACCAAGAAATACGACCTGAGCTGATAACATCATGTTCTCCCAATGTTCAGT 8701
 QY 8714 CGCGCAGATGCATCTGGCAAAAGGTTATCTACCTCACCGGTGACCCCAACCCGCCCT 8773
 Db 8702 CGCGCAGATGCATCTGGCAAAAGGTTATCTACCTCACCGGTGACCCCAACCCGCCCT 8761
 QY 8774 TGCACGCGCTGCGTGGGAGACAGTACACACTCCAAATCAACTTGTGCTTGGCTTGGCAATAT 8833
 Db 8762 CGCACGGGCTGCGTGGGAGACAGTACACACTCCAAATCAACTTGTGCTTGGCTTGGCAATAT 8821
 QY 8834 CATCATGTATGCGCCCAACCTATGGCAGAGATGATTTCTGATGACTCACTTTTCTCCAT 8893
 Db 8822 CATCATGTATGCGCCCAACCTATGGCAGAGATGATTTCTGATGACTCACTTTTCTCCAT 8881
 QY 8894 CTTTCTAGCTCAGAGCAACTTGAAGCCCTGGATTTGTCAGATCTACGGGCTTGTCTA 8953
 Db 8882 CTTTCTAGCTCAGAGCAACTTGAAGCCCTGGATTTGTCAGATCTACGGGCTTGTCTA 8941
 QY 8954 CTCATATGACCACTTGACCTACTCATGATCAATGTAACGACTCCATGGTCTTAGCCAT 9013
 Db 8942 CTCATATGACCACTTGACCTACTCATGATCAATGTAACGACTCCATGGTCTTAGCCAT 9001
 QY 9014 TACACTCCAGTATCTTCCAGGTGAGATCAATAGGTTGGCTTCATGCTCAGGAACT 9073
 Db 9002 TTAACCTCCAGTATCTTCCAGGTGAGATCAATAGGTTGGCTTCATGCTCAGGAACT 9061

QY 9074 TGGGTACCAACCTTGGAACTTGGAGCATCGGCGCAGAGTGTCCGCGCTAAGCTACT 9133
 Db 9062 TGGGTACCAACCTTGGAGTCTGGAGCATCGGCGCAGAGTGTCCGCGCTAAGCTACT 9121
 QY 9134 GTCCAGGGGGGAGGGCGGCACTTGTGGCAGATACCTTTTAACCTGGGCACTAAGGAC 9193
 Db 9122 GTCCAGGGGGGAGGGCTGCCACTTTCGGCAGTACTCTTTCACTGGGCACTAAGGAC 9181
 QY 9194 CAAGCTTAACTCACTCAATPCCCGCGGTCCTCCAGCTGGACTGTCTGGCTGGTTCGT 9253
 Db 9182 CAAGCTTAACTCACTCAATPCCCGGTCCTCCAGCTAGACTGTTCGGCTGGTTCGT 9241
 QY 9254 CGCTGGTTACAGCGGGGAGACATATATCACAGCTGTCTGTCGCCGACCCCGCTGGTT 9313
 Db 9242 TGCTGGTTACAGCGGGGAGACATATATCACAGCTGTCTGTCGCCGACCCCGCTGGTT 9301
 QY 9314 TCGTTCGTGCTACTCTCTACTTTTCTGTAGGGTAGGCATTTACCTGTCCTCCCAACCGATG 9373
 Db 9302 CATGTTGTGCTACTCTCTACTTTTCTGTAGGGTAGGCATTTACCTGTCCTCCCAACCGATG 9361
 QY 9374 AACGGGAGCTAACCACTCCAGGCTTAAGCCATTTCTCTTTTCTTTT 9424
 Db 9362 AACGGGAGCTAACCACTCCAGGCTTAAGCCATTTCTCTTTTCTTTT 9412

RESULT 9
 AAQ80498
 ID AAQ80498 standard; DNA; 9413 BP.
 AC AAQ80498;
 XX
 XX 16-OCT-2003 (revised)
 DT 06-SEP-1995 (first entry)
 XX
 DE DNA encoding HCV protein cleavable with new serine proteinase.
 XX
 KW proteinase; serine; cleavage; hepatitis C virus; HCV; ss.
 XX
 OS Hepatitis C virus; Virus.
 XX
 FH Key Location/Qualifiers
 FT CDS 330..9362
 FT /*tag= a
 FT /note= "HCV protein"
 XX
 JP06315377-A.
 XX
 XX 15-NOV-1994.
 PD
 XX 06-MAY-1993; 93JP-00105666.
 PF
 XX 06-MAY-1993; 93JP-00105666.
 PR
 XX {KAEN/} KAENNO K.
 PA {SUMQ } SUMITOMO METAL IND LTD.
 PA {SOYA- } SOYAKU GIJUTSU KENKYUSHO KK.
 XX
 WPI; 1995-032330/05.
 DR P-PSDB; AAR68622.
 XX
 PT New HCV-originated proteinase active substance - used for site-specific
 PT cleavage by an intermolecular reaction and the purification thereof.
 XX
 PS Disclosure; Page 10-19; 23pp; Japanese.
 CC This nucleotide encodes a protein from HCV (hepatitis C virus) which is
 CC cleaved between amino acids 2419 and 2420, by a new serine protease.
 CC contg. the sequence of AAR68621. The proteinase is purified as a fused
 CC product with the dihydrofolate reductase protein by using a methotrexate
 CC column. It can be used for the development of an inhibitor for HCV
 CC proteinase. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 9413 BP; 1887 A; 2818 C; 2690 G; 2018 T; 0 U; 0 Other;

Query Match			
Best Local Similarity 92.6%; Pred. No. 0;			
Matches 8713; Conservative			
QY	14	TGGGGGCGACACCTCCACCATGAATCACTCCCTGTTGAGGAACCTACCTGTTCTTCACGAGAA	73
DB	2	TGGGGGCGACACCTCCACCATGAATCACTCCCTGTTGAGGAACCTACCTGTTCTTCACGAGAA	61
QY	74	AGCGTCTAGCCATGCGGCTTATGATGCTGTCAGCCCTCCAGGACCCGCCCTCCCGGG	133
DB	62	AGCGTCTAGCCATGCGGCTTATGATGCTGTCAGCCCTCCAGGACCCGCCCTCCCGGG	121
QY	134	AGAGCCATAGTGTCTCGGAAACCGGTGACTAACCGGAATCCAGGACGACCGGGTCC	193
DB	122	AGAGCCATAGTGTCTCGGAAACCGGTGACTAACCGGAATCCAGGACGACCGGGTCC	181
QY	194	TTTCTTGGATCAACCCGCTCAATGCTCGAGATTTGGCGGTGCCCCGCGAGACTGCTAG	253
DB	182	TTTCTTGGATCAACCCGCTCAATGCTCGAGATTTGGCGGTGCCCCGCGAGACTGCTAG	241
QY	254	CCGAGTATGTTGGTTCGGAAAGGCTTTGTGTAATGCTGCTGATAGGCTGCTTGGAGTG	313
DB	242	CCGAGTATGTTGGTTCGGAAAGGCTTTGTGTAATGCTGCTGATAGGCTGCTTGGAGTG	301
QY	314	CCCGGAGGTCTCTGTAGACCGTGCACATGAGCAGCAATCTTAACCTCAAGAAAAAC	373
DB	302	CCCGGAGGTCTCTGTAGACCGTGCACATGAGCAGCAATCTTAACCTCAAGAAAAAC	361
QY	374	CAAACTTAACAAACCGCGCCCAAGGACGTCAGTTCCTTAACCTCAAGAAAAAC	433
DB	362	CAAACTTAACAAACCGCGCCCAAGGACGTCAGTTCCTTAACCTCAAGAAAAAC	421
QY	434	TGGTGAATTTACTGTTCCCGGCGAGGGGCCCGAGTTGGGTGCGCGGCACTAGGAA	493
DB	422	TGGTGAATTTACTGTTCCCGGCGAGGGGCCCGAGTTGGGTGCGCGGCACTAGGAA	481
QY	494	GACTTCCGAGCGTCCGACCTGCTGGAAGGCAACCTATCCCAAAGGCTCGCGACC	553
DB	482	GACTTCCGAGCGTCCGACCTGCTGGAAGGCAACCTATCCCAAAGGCTCGCGACC	541
QY	554	CGAGGCGAGGGCTCGGGCTCAGCCCGGTACCCCTTGGCCCTCTATGGCAATGAGGGCT	613
DB	542	CGAGGCGAGGGCTCGGGCTCAGCCCGGTACCCCTTGGCCCTCTATGGCAATGAGGGCT	601
QY	614	GGGTGGGCGAGGATGGCTCTGTCAACCCCGGCTCCCGGCTAGTTGGGGCCCAACGA	673
DB	602	GGGTGGGCGAGGATGGCTCTGTCAACCCCGGCTCCCGGCTAGTTGGGGCCCAACGA	661
QY	674	CCCCCGGCTAGGCTCGGTAACTTGGGTAAAGTCAATGATACCTTAATGCGGCTTCCG	733
DB	662	CCCCCGGCTAGGCTCGGTAACTTGGGTAAAGTCAATGATACCTTAATGCGGCTTCCG	721
QY	734	CGATCTCATGGGGTCAATTCGCTCGCGGCCCTTGGGGCGGCTGCGAGGCTT	793
DB	722	CGATCTCATGGGGTCAATTCGCTCGCGGCCCTTGGGGCGGCTGCGAGGCTT	781
QY	794	GGCACACGCTGTCGGGTCTGGAGGACGGGTGAATATGATACCTTAATGCGGCTTCCG	853
DB	782	GGCACACGCTGTCGGGTCTGGAGGACGGGTGAATATGATACCTTAATGCGGCTTCCG	841
QY	854	TTGCTCTTCTCTATCTTCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	913
DB	842	TTGCTCTTCTCTATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	901
QY	914	TTATGAATGGGCAACGCTTCGGGATATACATGTACGAAACGACTGCTCCAACTCAAG	973
DB	902	TTATGAATGGGCAACGCTTCGGGATATACATGTACGAAACGACTGCTCCAACTCAAG	961
QY	974	CATTGTGTATGAGGCGAGGACGTAATGATGATCTCCGGGTGCTGCTGCTGCTGCTGCT	1033
DB	962	TATTGTGTATGAGGCGAGGACATGATCATGACACACCCCGGGTGGCTGCTGCTGCTGCT	1021

QY	1034	GGAGGTTAACAAGCTCCCGTTCGGTGGTACGCTCACTCCACGCTCGCGGCGAGGAATGC	1093
DB	1022	GGAGGTTAATTTCTCCCGTTCGGTGGTACGCTCACTCCACGCTCGCGGCGAGGAACAG	1081
QY	1094	CAGCGTCCCCACTAGCAAAATACGACCGCACCTGCTGCTTGGGACGGCTGCTTT	1153
DB	1082	CAGCATCCCCACCAAGCAATACGACCGCACCTGCTGCTTGGGACGGCTGCTTT	1141
QY	1154	CTGCTCCGCTATGTAGTGGGGATCTCTGCGGATCTATTTCTCTGCTTCCAGCTGTT	1213
DB	1142	CTGTTCCGCTATGTAGTGGGGATCTCTGCGGATCTCTGCTTCCAGCTGTT	1201
QY	1214	CACCTTCTCGCTCCCGCATGAGACAGTGCAGCACTGCACTGCTCAATCTATCCCGG	1273
DB	1202	CACCTTCTCACTCCCGCATGAGACAGTGCAGCACTGCACTGCTCAATCTATCCCGG	1261
QY	1274	CCATGATACAGTCAACGCTATGGCTTGGATATGATGATGATGATGATGATGATGATGAT	1333
DB	1262	CCACGATACAGTCAACGCTATGGCTTGGATATGATGATGATGATGATGATGATGATGAT	1321
QY	1334	CCTAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1393
DB	1322	CCTAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1381
QY	1394	CCACTGGGAGTCTCTGGCGGCTTGGCTTACTTATTCATGTTAGGAACTGGGCTTAAGT	1453
DB	1382	CCACTGGGAGTCTCTGGCGGCTTGGCTTACTTATTCATGTTAGGAACTGGGCTTAAGT	1441
QY	1454	TCTGATTTGGCGCTTACTCTTTTGGCGGCTTGAACGGGAGACCCACACGACGGGAGGT	1513
DB	1442	CTTGAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1501
QY	1514	GSCCGGCGACACCCACTCCGGGTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1573
DB	1502	GSTAGCTCCAGACCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1561
QY	1574	CCAGTCTGTGAATACCAACGCGAGCTGGCAATCAACAGGACTGCTGCTGCTGCTGCTGCTGCT	1633
DB	1562	CCAACTCGTGAACACCAACGCGAGCTGGCAATCAACAGGACTGCTGCTGCTGCTGCTGCTGCT	1621
QY	1634	CTCCTCTCAAACCTGGTCTTGTTCGGGCTGTTTACGACACAAAGTTCACACTGCTCGG	1693
DB	1622	CTCCTCTCAAACCTGGTCTTGTTCGGGCTGTTTACGACACAAAGTTCACACTGCTCGG	1681
QY	1694	GTGCGCGGAGCGCATGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1753
DB	1682	GTGCGCGGAGCGCATGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1741
QY	1754	CATCACTTATTAAGCTTAAACGCTCGGATCAAGAGGCTTATTCGCTGCTGCTGCTGCTGCTGCT	1813
DB	1742	CATCACTTATTAAGCTTAAACGCTCGGATCAAGAGGCTTATTCGCTGCTGCTGCTGCTGCTGCT	1801
QY	1814	TGACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1873
DB	1802	TGACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1861
QY	1874	AAGCCCTTGTGTGTGGGACCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1933
DB	1862	GAGCCCTTGTGTGTGGGACCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1921
QY	1934	GAATGAGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1993
DB	1922	GAATGAGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1981
QY	1994	CTGTACATGATGAATAGTACTGGTTCATTAAGCTGCGGAGGTCCCGCTGCTGCTGCTGCTGCT	2053
DB	1982	GTGCACTGATGAACAGCACTGGGTTCACCAAGCTGCGGAGGTCCCGCTGCTGCTGCTGCTGCT	2041
QY	2054	CGGGGGGTTCGGTAAACCGACCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2113
DB	2042	CGGGGGGTTCGGTAAACCGACCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2101
QY	2114	GGCTACTTACACAAAATGTGGCTCGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2173

Db 2102 GGCACCTTACAAAGTGTGGCTGGGGCCCTGGTTGACACCCAGGTGCATGGTTGACTA 2161
QY 2174 CCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTCCATCTTTAAGGTTAGCAT 2233
Db 2162 CCATACAGGCTCTGGCACTACCCCTGCACTCTCAATTTTCCATCTTTAAGGTTAGCAT 2221
QY 2234 GTATGTGGGGGGCTGGAGCAAGGCTCAATCCGCACTGCAATTTGAACTCGAGAGAGCG 2293
Db 2222 GTATGTGGGGGGCTGGAGCAAGGCTCAATCCGCACTGCAATTTGAACTCGAGAGAGCG 2281
QY 2294 CTGTAACTTGGAGGACAGGATAGGTGAGAACTCAGCCCGCTGCTGTCTGTCTAACAAGA 2353
Db 2282 CTGTAACTTGGAGGACAGGATAGGTGAGAACTCAGCCCGCTGCTGTCTGTCTAACAAGA 2341
QY 2354 GTGGCAGATACTGCCCTCTGTCTTCAACACCCCTACCGGCTTTATCCATCTGTCTTATCCCA 2413
Db 2342 GTGGCAGATACTGCCCTCTGTCTTCAACACCCCTACCGGCTTTATCCCACTGGCTTATCCCA 2401
QY 2414 TCTCATCAGAACATCGTGGAGCGTGCAATACTGTACGCTGTAGGGTCAGGCTTTGTCTC 2473
Db 2402 TCTTCAACGGAACATCGTGGAGCGTGCAATACTGTACGCTGTAGGGTCAGGCTTTGTCTC 2461
QY 2474 CTTTTGCATCAAAATGGGAGTACNTCTCTGCTTCTCTTCTCTCTCTCTCTCTCTCTCTCT 2533
Db 2462 CTTTTGCATCAAAATGGGAGTATATCTCTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2521
QY 2534 GTGTGCTCTGTCTGTGTGATGCTCTGATAGCCACGAGCTCAGGCGGCTTTAGAGAACTT 2593
Db 2522 CTGTGCTCTGTCTGTGTGATGCTCTGATAGCCACGAGCTCAGGCGGCTTTAGAGAACTT 2581
QY 2594 GTGTGCTCTGTGTGATGCTCTGATAGCCACGAGCTCAGGCGGCTTTAGAGAACTT 2653
Db 2582 GTGTGCTCTGTGTGATGCTCTGATAGCCACGAGCTCAGGCGGCTTTAGAGAACTT 2641
QY 2654 CTTCCTGCGCGCTGTGTGATGCTCTGATAGCCACGAGCTCAGGCGGCTTTAGAGAACTT 2713
Db 2642 CTTCCTGCGCGCTGTGTGATGCTCTGATAGCCACGAGCTCAGGCGGCTTTAGAGAACTT 2701
QY 2714 TGGCGTATGSCGCTGTCTCTGCTCTCTGCTGCTTACCAACGAGCTTACGCTTGGGA 2773
Db 2702 TGGCGTATGSCGCTGTCTCTGCTCTCTGCTGCTTACCAACGAGCTTATGCCATGGA 2761
QY 2774 CCGGGAATGCTGCTCATCTGTGGGGGTGCGGTCTTGTAGGCTGTGTAATCTTGAACCTT 2833
Db 2762 CCGAGATAGGCTGTGCTCATCTGTGGAGCGCGGTCTTGTAGGCTGTGTAATCTTGAACCTT 2821
QY 2834 GTCAACCATCTACAAAGTGTCTCTACTAGGCTCATATGTTGTTTGAATCAATATTTATCAC 2893
Db 2822 GTCAACCATCTATAGGTTGTCTCTGCTAGGCTCATATGTTGTTTGAATCAATATTTATCAC 2881
QY 2894 CAGAGCGAGGCGCATATGAAAGTGTGGTCCCGCCCTCAAGCTTCGGGAGGCGCGGA 2953
Db 2882 CAGAGCGAGGCGCATATGAAAGTGTGGTCCCGCCCTCAAGCTTCGGGAGGCGCGGA 2941
QY 2954 TGCATCATCTCTCTCATCTGTGGGTCTATCAGAGTTAATTTTGAATCAATCAACAACT 3013
Db 2942 TGCATCATCTCTCTCATCTGTGGGTCTATCAGAGTTAATTTTGAATCAATCAACAACT 3001
QY 3014 CTGTCTCGGCATCTCGGCGCGCTCATGTGTCTCCAGGCTGCGATACAGAGAGTGGCTA 3073
Db 3002 CTGTCTCGGCATCTCGGCGCGCTCATGTGTCTCCAGGCTGCGATACAGAGAGTGGCTA 3061
QY 3074 CTGTCTCGGCTCAAGGCTCATCTGTGATGATGTTAGTGGAAAGTCCCGGGGG 3133
Db 3062 CTGTGTACGGCTCAGGGCTCATCTGTGTGATGATGTTAGTGGGAGAGTCTCGTGGAGG 3121
QY 3134 TCATTTATGTCATATGTTCTCATGTAGTGTGGGGGCTGACAGGTACGTACGTTTATAA 3193
Db 3122 CCATATGTCATATGTTCTCATGTAGTGTGGGGGCTGACAGGTACGTACGTTATGA 3181
QY 3194 CCATCTTACCCACTCGGAGCTGGGCCACGCGGCTTACGAGACCTTGGCGTGGCGT 3253

Db 3182 CCATCTTACTCCATGCTGGGATTTGGGCCCAACGCGGCTACGAGACCTTGCCTGGGAGT 3241
QY 3254 AGAGCCGCTGTCTCTTCCGCCATGGAGACCAAGGTATATCACTGGGGAGCAGACACCGC 3313
Db 3242 AGAGCCGCTGTCTCTTCCGCCATGGAGACTAACTCATCACTGGGGAGCAGACACCGC 3301
QY 3314 TGCCTGTGGGACATCATCTTGGCTTACCGCTCTCCGCCCGGAGGGGAGGAGATATT 3373
Db 3302 GGCCTGTGGGACATCATCTTGGCTTACCGCTCTCCGCCCGGAGGGGAGGAGATATT 3361
QY 3374 TTTGGGACCGGCTGATAGTCTCGAAGGGCAAGGCTGGCGACTCTTTCGCCCATCAACGGC 3433
Db 3362 TCTAGGACCGGCTGATAGTCTTGGAGAGCAGGGGTGCGGCTCTTTCGCCCATCAACGGC 3421
QY 3434 CTACTCCCAACAACCGGGGCTTACTTGTGTGATCATCACTAGCTCTCAAGSCCGGA 3493
Db 3422 CTATTCCTCAACAAACGCGGGGCTTGTGTGTGCTGTATCATCACTAGCTCTCAAGSCCGGA 3481
QY 3494 CAAGAACAGGTGCAAGGGGAGGTTCAAGTGTCTTACCGCAACACAANTCTTCTCTGGC 3553
Db 3482 CAAGAACAGGTGCAAGGGGAGGTTCAAGTGTCTTACCGCAACACAANTCTTCTCTGGC 3541
QY 3554 GACCTGCATCAACCGGCTGTGTGTGACTGTCTACCATGGCGCTGGTTCGAGAACCTTAGC 3613
Db 3542 GACCTGTCTCAATGGCGTGTGTGTGAGCCGCTTACCATGGTGTGCGGCTCGAAGACCTTAGC 3601
QY 3614 CGGTTCGAAAGGTCCATCAACCAATGTACCAATGTAGACCTGAGACCTGTGGCTG 3673
Db 3602 CGGCGCGGAGGGTCCATATCAACCAATGTAGACCTGAGACCTGTGGCTG 3661
QY 3674 GCAGGCGCGGCGGCGCTCCATGACACCATGACAGCTGTGGAGCTTCGACCTTTA 3733
Db 3662 GCAGGCGCGGCGGCGGCTCCATGACACCGTGTGACACCTGTGGAGCTTCGACCTTTA 3721
QY 3734 CTGTGTACAGAGACATGTCTGATGTCTATTCGCTGCGGCGCGGCGGCGGAG 3793
Db 3722 CTGTGTACAGAGACATGTCTGATGTCTATTCGCTGCGGCGGCGGCGGAG 3781
QY 3794 TCTACTCTCCCGGCGGCTCTCTTACCTGAAAGGCTCTCGGCTGTGCTTGTG 3853
Db 3782 CTGTGTACAGAGACATGTCTGATGTCTATTCGCTGCGGCGGCGGCGGAG 3841
QY 3854 CCCTTCGCGGCGGCGGCTCTTCCGCGGCTGTGTGACCGCGGCGGCTTCGCGAA 3913
Db 3842 CCCTTCGCGGCGGCGGCTGTGTGAGGCTCTTCCGCGGCTGTGTGACCGCGGCGGCTTCGCGAA 3901
QY 3914 GCGGTGTGACTTCATACCGCTGTGAGTCTATGGAACCTACCATGCGGCTCTCGGCTTTCAC 3973
Db 3902 GCGGTGTGACTTCATACCGCTGTGAGTCTATGGAACCTACCATGCGGCTCTCGGCTTTCAC 3961
QY 3974 AGACAACTCAACCCCGGCTTACCGGAGCATTCGAAAGTGGGACATCTGACGCTCC 4033
Db 3962 AGACAACTCATCCCTTCGCGGCGTACCGGAACTTCCAAAGTGGGACATTCACGCTCC 4021
QY 4034 TACTGACGCGGCAAGAGACCAAAAGTGGCGGCTGTGTGACGCGGCGGCTTCAAGGT 4093
Db 4022 CACTGACGCGGCAAGAGACCAAAAGTGGCGGCTGTGTGACGCGGCGGCTTCAAGGT 4081
QY 4094 GCTGTCTGTAAACCGCTCGGCTGCGGCACTTACGCTTGGGGGTATATCTTCAAGGC 4153
Db 4082 GCTGTCTGTAAACCGCTCGGCTGCGGCACTTACGCTTGGGGGTATATCTTCAAGGC 4141
QY 4154 ACACGATTCGACCCCTTACATCAGAACTGGGTGAGGACCATTAACGCGGCGGCTCCAT 4213
Db 4142 ACATGACATTCGAGGCTTACATCAGAACTGGGTGAGGACCATTAACGCGGCGGCTCCAT 4201
QY 4214 TACTGACTCCACTATGGCAAGTCTCTTCCGCAAGTGTGTGCGGCGGCTTATGA 4273
Db 4202 CAGGTACTCCACTATGGCAAGTCTCTTCCGCAAGTGTGTGCGGCGGCTTATGA 4261
QY 4274 CATCATATATGTGATGATGCCACTCACTGACTCGACTCATCTTTGGGATCGGAC 4333
Db 4262 CATCATATATGTGATGATGCCACTCACTGACTCGACTCATCTTTGGGATCGGAC 4321

QY 4334 AGTCCTGGACCAACGGAGACGGCTGGAGCGCGCTCGTCGTGCTGCGCCACCGCTACAC 4393
DB 4322 AGTCCTGGATCAGGCAGACAGCGCTGGAGCGCGCTCGTCGTGCTGCGCCACCGCC 4381
QY 4394 TCCGGGATCCGTTACCGTGCACACCCCAATATCGAGGAATAGCGCTGTCCAAATGG 4453
DB 4382 TCCGGGATCCGATCCCGTGCACACCCCAATATCGAGGAATAGCGCTGTCCAAATGG 4441
QY 4454 AGAGATCCCTTTATGGCAAGACATCCCAATTGAGGCCATCAAGGGGGGAGGCATCT 4513
DB 4442 AGAGATCCCTTTATGGCAAGACATCCCAATTGAGGCCATCAAGGGGGGAGGCATCT 4501
QY 4514 CATTTTCTGCCATTCOAAGAAAGAAATGTGACAGCTCGCGCAAAAGCTGACAGCCCTCG 4573
DB 4502 CATTTTCTGCCATTCOAAGAAAGAAATGTGACAGCTCGCGCAAAAGCTGACAGCCCTCG 4561
QY 4574 ACTGAAACGCTGTAGCATATTAACGGGGCTTTGANGTGTCCGTATACAGCGCTATCGGAGA 4633
DB 4562 ACTCAATGCTGTAGCTATTAACGGGGCTTCGATGTGTCCGTATACAGCGCTATCGGAGA 4621
QY 4634 CGTCGTTGCTGGCAACAGACGCTCTAATGACGGGTTTACCGGCGAATTTGACTCAGT 4693
DB 4622 CGTCGTTGCTGGCAACAGACGCTCTAATGACGGGTTTACCGGCGAATTTGACTCAGT 4681
QY 4694 GATCGACTGCANTACATGTGTACCCCAAGACAGTGCATCTCAGCTTGGATCCCACTTCAC 4753
DB 4682 GATCGACTGCANTACATGTGTACCCCAAGACAGTGCATCTCAGCTTGGATCCCACTTCAC 4741
QY 4754 CATTGAGACGACACCGTGCCTCCCAAGACGCGGTGTCCGCTCCGCAACGGCGAGGTAGAAC 4813
DB 4742 CATTGAGACGACACCGTGCCTCCCAAGACGCGGTGTCCGCTCCGCAACGGCGAGGTAGAAC 4801
QY 4814 TGGCAGGGTAGAGTGGCATCTACAGTTTGTGACTCCAGGAAACGGCCCTCGGGCAT 4873
DB 4802 TGGCAGGGTAGAGTGGCATCTACAGTTTGTGACTCCAGGAAACGGCCCTCGGGCAT 4861
QY 4874 GTTCGATTCCTCGTCTGTGCTAGTGTATGACGCGGGCTGTGCTTGTGTATAGTTCAC 4933
DB 4862 GTTCGATTCCTCGTCTGTGCTAGTGTATGACGCGGGCTGTGCTTGTGTATAGTTCAC 4921
QY 4934 GCCCGCTGAGACCTCGGTAGGTTGGGGTGTACCTAAATACACCGGTTGCCGCTG 4993
DB 4922 GCCCGCTGAGACCTCGGTAGGTTGGGGTGTACCTAAATACACCGGTTGCCGCTG 4981
QY 4994 CAGGACCACTGGAGTGTGGGAGAGCGTCTTCAAGGCTTCAACCAACATAGATGCCCA 5053
DB 4982 CAGGACCACTGGAGTGTGGGAGAGCGTCTTCAAGGCTTCAACCAACATAGATGCCCA 5041
QY 5054 CTTCCTGTCCAGACTAAACAGGACGAGACACTTCTTACAGGCTTCAACCAACATAGATGCCCA 5113
DB 5042 CTTCCTGTCCAGACTAAACAGGACGAGACACTTCTTACAGGCTTCAACCAACATAGATGCCCA 5101
QY 5114 TACAGTGTGGCGAGGCTCAAGCTCCACCTCCATGTTGGGACCAAAATGTGAAGTGTCT 5173
DB 5102 TACAGTGTGGCGAGGCTCAAGCTCCACCTCCATGTTGGGACCAAAATGTGAAGTGTCT 5161
QY 5174 CATACGCTGAACTACACTGACAGGCGCAACACCCCTGCTCTATAGGCTAGGACCGT 5233
DB 5162 CATACGCTGAACTACACTGACAGGCGCAACACCCCTGCTCTATAGGCTAGGACCGT 5221
QY 5234 CCAAAATGAGGCTCATCTCCACACACCCCAATCACTAATACATGCGATGATGTGCG 5293
DB 5222 TCAAAATGAGGCTCATCTCCACACACCCCAATCACTAATACATGCGATGATGTGCG 5281
QY 5294 TGACCTGGAGGTGCTCACTAGACACCTGGTGTGGTAGGCGAGTCTTTCAGCTTTGCG 5353
DB 5282 TGACCTGGAGGTGCTCACTAGACACCTGGTGTGGTAGGCGAGTCTTTCAGCTTTGCG 5341
QY 5354 CGCATACTGCTTGACACAGGCGATGTGTCTATTTGGGCGAGATCATCTTGTCCGGAA 5413
DB 5342 CGCATACTGCTTGACACAGGCGATGTGTCTATTTGGGCGAGATCATCTTGTCCGGAG 5401

QY 5414 GCCAGCTGTGTTCCCGACAGGAAGTCTCTACAGGAGTTCGATGAGATGGAAGTG 5473
DB 5402 GCCAGCTGTGTTCCCGACAGGAAGTCTCTACAGGAGTTCGATGAGATGGAAGTG 5461
QY 5474 TGCTTCAACACTTCTTATATCAGCAGGGAATGCGAGCTGCGCGAGCAATTCAGCAAAA 5533
DB 5462 TGCTTCAACACTTCTTATATCAGCAGGGAATGCGAGCTGCGCGAGCAATTCAGCAAAA 5521
QY 5534 GGGCTTCGGGTTGTCGAACCGCCACCAAGCAGCAGGAGGCTGCTGCTCCCGTGTGGA 5593
DB 5522 GGGCTTCGGGTTGTCGAACCGCCACCAAGCAGCAGGAGGCTGCTGCTCCCGTGTGGA 5581
QY 5594 GTCCAAAGTCGCGAGCCCTTGAGACCTTCTGGGCGAAAGACATGTGGAATTCATACGCG 5653
DB 5582 GTCCAAAGTCGCGAGCCCTTGAGGCTTCTTCTGGGCGAAAGACATGTGGAATTCATACGCG 5641
QY 5654 AATACAGTACTAGCAGGCTTATCCACTCTGCTGTGGAACCCCGCGATAGCATTCAT 5713
DB 5642 AATACAGTACTAGCAGGCTTATCCACTCTGCTGTGGAACCCCGCGATAGCATTCAT 5701
QY 5714 GGCATTTACAGCTTCTATACACTAGCCGCTCACCACCAAAACACCCCTCTCTTAAAT 5773
DB 5702 GGCATTTACAGCTTCTATACAGCCGCTCACCACCAAAATACCCCTCTCTTAAAT 5761
QY 5774 CTGGGGGATGGTGGCTGCGCCAACTCGCTCTCTCCAGCGCTCGCTCAGCTTTCGTGG 5833
DB 5762 CTGGGGGATGGTGGCTGCGCCAACTCGCTCTCTCCAGCGCTCGCTCAGCTTTCGTGG 5821
QY 5834 CGCGGCATCGCGGAGCGCTGTTGGCAGCATAGCCCTTGGGAAGGTGCTCGTGACAT 5893
DB 5822 CGCGGCATTCGCGGCTGCGCGCTTGGCAGCATAGCTTTCGGGAAGGTGCTCGTGACAT 5881
QY 5894 CTGGCGGCTATGGGGCAGGGGTAGCCGCGCACTCGTGGCCCTTAAAGTTCATGAGCG 5953
DB 5882 TCTGGCGGCTATGGGGCAGGGGTAGCCGCGCACTCGTGGCCCTTAAAGTTCATGAGCG 5941
QY 5954 CAGAGTGCCCTCCACCGAGGACCTTGTCAACTTACTCCCTGCCATCTCTCTCTGGTGC 6013
DB 5942 CAGAGTGCCCTCCACCGAGGACCTTGTCAACTTACTCCCTGCCATCTCTCTCTGGTGC 6001
QY 6014 CTGTGCTCGGGTGTGTGCGCAGCAATATCTGCTGTGGCACTGCGGCGCGGAGAGG 6073
DB 6002 CTGTGCTCGGGTGTGTGCGCAGCAATATCTGCTGTGGCACTGCGGCGCGGAGAGG 6061
QY 6074 GGCCTGCAGTGTGAAGAACCGGCTCATAGCCTTCCCTTCCGGGGTAAACGCTCTCCC 6133
DB 6062 GGCCTGCAGTGTGAAGAACCGGCTCATAGCCTTCCCTTCCGGGGTAAACGCTCTCCC 6121
QY 6134 TACGCACTATGTGCTGAGCAGCAGCTGTCAGCAGCTGTCTCATCTCAGATCTCTTAGCCT 6193
DB 6122 TACGCACTATGTGCTGAGCAGCAGCAGCTGTCAGCAGCTGTCTCATCTCAGATCTCTTAGCCT 6181
QY 6194 TACCATCTCACTGCTGAGCGGCTCCACAGTGGATTAATGAGGACTGCTTACGCG 6253
DB 6182 TACCATCTCACTGCTGAGCGGCTCCACAGTGGATTAATGAGGACTGCTTACGCG 6241
QY 6254 ATGCTCCGCTCGTGGCTTAAAGGATGTTGGGATTTGGATATGCAACGCTGTGACGCTT 6313
DB 6242 ATGCTCCGCTCGTGGCTTAAAGGATGTTGGGATTTGGATATGCAACGCTGTGACGCTT 6301
QY 6314 CAGACCTGGCTCGAGTCCAACTCTCTCGCGGTTACCGGAGTCTCTTCTCTGTCTATG 6373
DB 6302 CAGACCTGGCTCGAGTCCAACTCTCTCGCGGTTACCGGAGTCTCTTCTCTGTCTATG 6361
QY 6374 CCAACGCGGGTACAGGGAGTCTGCGGGGGAAGGCAATCATGCAAAACCACTGCGGCTATG 6433
DB 6362 CCAACGCGGGTACAGGGAGTCTGCGGGGGAAGGCAATCATGCAAAACCACTGCGGCTATG 6421
QY 6434 CCGGACACAGATCGCGGACATGTCAAAACGCTTCCATGAGGATCGTAGGGCTTAGAAC 6493
DB 6422 TGGAGCACAGATCACCGGACATGTCAAAACGCTTCCATGAGGATCGTAGGGCTTAGAAC 6481
QY 6494 CTGACGCAACGCTGGCAGGAAAGTTCCTCCCATCAACGCAATACACGAGGACCTTGAC 6553

Db 6482 CTGACGAAACAGTGGGATGAAATCCCCATCAACGATACACACGGGCGCGTGAC 6541
 Qy 6554 ACCCTCCCGCGCGCCAACTATCCAGGGCGCTATGCGGGTGGCTGTGAGGAGTAGT 6613
 Db 6542 GCCCTCCCGCGCGCGCCAACTATCCAGGGCGCTATGCGGGTGGCTGTGAGGAGTAGT 6601
 Qy 6614 GGAGTTACGGGTGGGGGATTTCCACTAGCTACGCGGCATGACCACTGACAAAGTAA 6673
 Db 6602 GGAGTTACGGGTGGGGGATTTCCACTAGCTACGCGGCATGACCACTGACAAAGTAA 6661
 Qy 6674 GTGCCCATCCAGGTTCCGGGCCCGGAATTTCAAGGAGTGTGAGGAGTGGCGTTGCA 6733
 Db 6662 ATGCCCATGCCAGTTCCAGGCCCGGAATTTCAAGGAGTGTGAGGAGTGGCGTTGCA 6721
 Qy 6734 CAGTACGCTCCGGCGTGCAGAACCTCTCTACGGGAGGAGCTACGTTCCAGGTCGGCT 6793
 Db 6722 CAGTATGCTCCAGTGTGCAGAACCTCTCTACGGGAGGAGCTACGTTCCAGGTCGGCT 6781
 Qy 6794 CAACCAATACCTGTGGGTGGCGAGCTCCCATGGGAGCCCGAACCAGCTAACAGTCT 6853
 Db 6782 CAACCAATACCTGTGGGTGGCGAGCTCCCATGGGAGCCCGAACCAGCTAACAGTCT 6841
 Qy 6854 TACTTCCATCTACCGATCCCTCCGATTTACAGAGAGAGCGCTAAGCGTAGGCTGGC 6913
 Db 6842 CACTTCCATCTCCGACCCCTCTCATATTACAGAGAGAGCGCTAAGCGTAGGCTGGC 6901
 Qy 6914 TAGAGGCTCTCCCGCTCTTTAGCCAGCTCATACAGTAGCAGTTGCTCGGCTCTTT 6973
 Db 6902 CAGGGGTCTCCCGCTCTTTAGCCAGCTCATACAGTAGCAGTTGCTCGGCTCTTT 6961
 Qy 6974 GAAGGACATGCACTACCAACATGACTCCCGGAGCGTGAACCTCATCCAGGCGCAACCT 7033
 Db 6962 GAAGGACATGCACTACCAACATGACTCCCGGAGCGTGAACCTCATCCAGGCGCAACCT 7021
 Qy 7034 CTCTGGCGGACAGATGGCGGAAATCATCTCGGTGGAGTCAAGAAATAGGTAGT 7093
 Db 7022 CCGTGGCGGAGAGATGGCGGAAATCATACCGGTGGAGTCAAGAAATAGGTAGT 7081
 Qy 7094 AATTCTGGAATCTTCGAAACCGTTACGCGGAGGGGATGAGGGAGATATCCGTCG 7153
 Db 7082 AATCTGGAATCTTCGAAACCGTTACGCGGAGGGGATGAGGGAGATATCCGTCG 7141
 Qy 7154 GGCGGAGATCTCGGAAATTCAGGAATTCCTTCAGCGTTGCCATATGGGACGCGCC 7213
 Db 7142 GGCGGAGATCTCGGAAATTCAGGAATTCCTTCAGCGTTGCCATATGGGACGCGCC 7201
 Qy 7214 GGAATCAATCTCCAGTGTAGAGTCTCGGAGGACCGGACTACGTCCTCGGTGGT 7273
 Db 7202 GGAATCAATCTCCAGTGTAGAGTCTCGGAGGATCCGACTACGTCCTCGGTGGT 7261
 Qy 7274 ACAGGATGCCATTCGCAACTACCAAGGCTCCCAATACCACTCCAGGAGAAAGAG 7333
 Db 7262 ACAGGATGCCATTCGCAACTACCAAGGCTCCCAATACCACTCCAGGAGAAAGAG 7321
 Qy 7334 GAGGTTGTCTCGAGAAATTCGAATGTCTTCGCTGGGAGGCTCGCCACTAAGAC 7393
 Db 7322 GAGGTTGTCTCGAGAAATTCGAATGTCTTCGCTGGGAGGCTCGCCACTAAGAC 7381
 Qy 7394 CTTCGTTAGCTCCGGATCGTGGCGCTTGATAGGGAACCGGACCGCTTCCTGACCT 7453
 Db 7382 CTTCGTTAGCTCCGGATCGTGGCGCTTGATAGGGAACCGGACCGCTTCCTGACCT 7441
 Qy 7454 GGCTCCGACGAGTGCAGAAAGATCCGACGTTGAGTGTCTCTCTCCATGCCCGCT 7513
 Db 7442 GGCTCCGACGAGTGCAGAAAGATCCGACGTTGAGTGTCTCTCTCCATGCCCGCT 7501
 Qy 7514 TGAAGGAGGAGCGGGAACCGGATCTCAGGACGCTTTGCTTACCGTGAAGTGA 7573
 Db 7502 CGAGGAGAGCGGGAACCGGATCTCAGGACGCTTTGCTTACCGTGAAGTGA 7561
 Qy 7574 GGCTAGTGAAGTGTCTGCTGCTCAATGTCTTATAGTGTGACAGGCGCTGATCAC 7633

Db 7562 AGCTGGTGAGGACGTCGCTCTGCTCAATGTCTTATACATGACAGCTGCTTGTATCAC 7621
 Qy 7634 GCAATGCTGGGAGGAAAGTATGTCGCCATCAACCCGTTGAGCAACTCTTTCTGCTGG 7693
 Db 7622 GCAATGCTGGGAGGAAAGTATGTCGCCATCAATCCGTTGAGCAACTCTTTCTGCTGG 7681
 Qy 7694 TCACCAAAATGCTGTCGACCAACATCCGAGCGCAAGCTCCGCGAGAGAGAGT 7753
 Db 7682 TCACCAAAATGCTGTCGACCAACATCCGAGCGCAAGCTCCGCGAGAGAGAGT 7741
 Qy 7754 CACTTTGACAGATTCGAGTCTCTGGATGATTAATCCGGGACGTTCTCAAGGAGATGAA 7813
 Db 7742 CACTTTGACAGATTCGAGTCTCTGGATGATTAATCCGGGACGTTCTCAAGGAGATGAA 7801
 Qy 7814 GCGGAGGCGCTCACAGTTAAGCTAAGCTTCTATCTATAGAGAGGCGCTGCAAGTGCAC 7873
 Db 7802 GCGGAGGCGCTCACAGTTAAGCTAAGCTTCTATCTATAGAGAGGCGCTGCAAGTGCAC 7861
 Qy 7874 GCGGAGGCGCTCACAGTTAAGCTAAGCTTCTATCTATAGAGAGGCGCTGCAAGTGCAC 7933
 Db 7862 GCGGAGGCGCTCACAGTTAAGCTAAGCTTCTATCTATAGAGAGGCGCTGCAAGTGCAC 7921
 Qy 7934 CAGGAGGCGCTTAACCACTCCGCTCCGCTGGGAGGACTTCTGCGAGACACTGAAAC 7993
 Db 7922 CAGGAGGCGCTTAACCACTCCGCTCCGCTGGGAGGACTTCTGCGAGACACTGAAAC 7981
 Qy 7994 ACCAATGTGACCAACCACTCATGSCAAAGTAGGTCTTCTGCTCCACACGAGAGGG 8053
 Db 7982 ACCAATGTGACCAACCACTCATGSCAAAGTAGGTCTTCTGCTCCACACGAGAGGG 8041
 Qy 8054 AGCGCGCAAGCGAGCTCGCTTATGCTATCCAGACCTGGGAGTCTGCTATGCCGAA 8113
 Db 8042 AGCGCGCAAGCGAGCTCGCTTATGCTATCCAGACCTGGGAGTCTGCTATGCCGAA 8101
 Qy 8114 GATGGCGCTTACGAGTGTCTCCACCTTCTCAGGCGCTGATGGGCTCTCATACGG 8173
 Db 8102 GATGGCGCTTACGAGTGTCTCCACCTTCTCAGGCGCTGATGGGCTCTCATACGG 8161
 Qy 8174 ATTTCAATCTCCCAAGAGCGGCTGAGTCTCTGCTGATTAATCCAGAA 8233
 Db 8162 ATTTCAATCTCCCAAGAGCGGCTGAGTCTCTGCTGATTAATCCAGAA 8221
 Qy 8234 ATGCCCTATGGGCTTCTCATATGACACCGCTGTTTGTACTCAACGCTCACTGAGAGTGA 8293
 Db 8222 ATGCCCTATGGGCTTCTCATATGACACCGCTGTTTGTACTCAACGCTCACTGAGAGTGA 8281
 Qy 8294 CATCTGTTGAGGAGTCAATTTACCAATGTTGTGACTTTGGCGCGGAGGCGGAGAGG 8353
 Db 8282 CATCTGTTGAGGAGTCAATTTACCAATGTTGTGACTTTGGCGCGGAGGCGGAGAGG 8341
 Qy 8354 CATAGGCTCGCTCACAGAGGCGCTTACATCGGGGTCCCTGACTAATCCAAAGGGA 8413
 Db 8342 CATAGGCTCGCTCACAGAGGCGCTTATGTCGGGGTCCCTGACTAATTCGAAAGGGA 8401
 Qy 8414 GAATCGGCTTATCGCGGTGCGGCAAGTGGCTGCTGACGACTAGTCTGAGTAAATAC 8473
 Db 8402 GAATCGGCTTATCGCGGTGCGGCAAGTGGCTGCTGACGACTAGTCTGAGTAAATAC 8461
 Qy 8474 CACTCATGTTACTTGAAGGCGACTGCGAGCTGTGAGCTGCAAGCTCCAGGAGTGCAC 8533
 Db 8462 CACTCATGTTACTTGAAGGCGACTGCGAGCTGTGAGCTGCAAGCTCCAGGAGTGCAC 8521
 Qy 8534 GATGTCGTGAGGAGGAGGAGCTTGTGCTGTTATCTGTGAAGCGGCGGAGACCGAGGGA 8593
 Db 8522 GATGTCGTGAGGAGGAGGAGCTTGTGCTGTTATCTGTGAAGCGGCGGAGACCGAGGGA 8581
 Qy 8594 TGCGGCGGCGCTACGAGCTTCAAGGAGGCTTGTGCTGTTATCTGTGAGAGTGGGAGACCGAGGGA 8653
 Db 8582 TGCGGCGGCGCTACGAGCTTCAAGGAGGCTTGTGCTGTTATCTGTGAGAGTGGGAGACCGAGGGA 8641
 Qy 8654 TGCGGCGGCGCTACGAGCTTCAAGGAGGCTTGTGCTGTTATCTGTGAGAGTGGGAGACCGAGGGA 8713
 Db 8642 TGCGGCGGCGCTACGAGCTTCAAGGAGGCTTGTGCTGTTATCTGTGAGAGTGGGAGACCGAGGGA 8701

QY 8714 CCGCAGCATGTCATCTGGCAAAAGGATATACCTCAACCTGAGCCCAACACCCCT 8773
 Db |||||
 QY 8702 CCGCAGCATGTCATCCGGCAAAAGGATATACCTCAACCTGAGCCCAACACCCCT 8761
 Db |||||
 QY 8774 TGCAGGGCTGCTGGGAGAGCTAGACACACTCAATCAACTCTTGCTAGGCAATAT 8833
 Db |||||
 QY 8762 CGCAGGGCTGCTGGGAGAGCTAGACACACTCAATCAACTCTTGCTAGGCAATAT 8821
 Db |||||
 QY 8834 CATCATGTATGGCCCAACCTATGGCAAGGATGTTGTGATGACTCACTTTTCTCCAT 8893
 Db |||||
 QY 8822 CATCATGTATGGCCCAACCTATGGCAAGGATGTTGTGATGACTCACTTTTCTCCAT 8881
 Db |||||
 QY 8894 CTTTCTAGCTCAAGAGCAACTTGAAAAAGCCCTGGAATGTCAGATCTACGGGGCTTGCTA 8953
 Db |||||
 QY 8882 CTTTCTAGCTCAAGAGCAACTTGAAAAAGCCCTGGAATGTCAGATCTACGGGGCTTGCTA 8941
 Db |||||
 QY 8954 CTCATTGAGCCACTTGACCTTACCTCAGATCAATGAAAGCTTCCATGCTTACGCGCAT 9013
 Db |||||
 QY 8942 CTCATTGAGCCACTTGACCTTACCTCAGATCAATGAAAGCTTCCATGCTTACGCGCAT 9001
 Db |||||
 QY 9014 TACACTCCACAGTTACTCTCCAGGTGAGATCAATAGGCTGGCTTCATGCTCAGGAACT 9073
 Db |||||
 QY 9002 TTTACTCCACAGTTACTCTCCAGGTGAGATCAATAGGCTGGCTTCATGCTCAGGAACT 9061
 Db |||||
 QY 9074 TGGGTACACCTTGCGAACTTGAGACATCGGCCAGAAAGTGTCCCGCTAAGCTACT 9133
 Db |||||
 QY 9062 TGGGTACCGCTTGTGGAGTCTGGAGACATCGGCCAGAAAGTGTCCCGCTAAGCTACT 9121
 Db |||||
 QY 9134 GTCCAGGGGGAGGGCGGCACCTTGGGAGATACCTTTTAACTGGGCAATGAGAC 9193
 Db |||||
 QY 9122 GTCCAGGGGGAGGGCGGCACCTTGGGAGATACCTTTTAACTGGGCAATGAGAC 9181
 Db |||||
 QY 9194 CAAGCTTAACTACTCTCAATCCCGGCGCTGCCAGCTGGACCTGTCTGCTGCTGCT 9253
 Db |||||
 QY 9182 CAAGCTTAACTACTCTCAATCCCGGCGCTGCCAGCTGGACCTGTCTGCTGCTGCT 9241
 Db |||||
 QY 9254 CGCTGGTTACAGCGGGGAGACATATATCAGAGCTGTCTGTCGCCCAACCCGCTGCT 9313
 Db |||||
 QY 9242 TGCTGGTTTACACGGGGGAGACATATATCAGAGCTGTCTGTCGCCCAACCCGCTGCT 9301
 Db |||||
 QY 9314 TCCGTTGTGCTACTCTCTACTTTCTGTAGGGGTAGGCAATTTACCTGCTCCCAACCGATG 9373
 Db |||||
 QY 9302 CATGTTGTGCTACTCTCTACTTTCTGTAGGGGTAGGCAATTTACCTGCTCCCAACCGATG 9361
 Db |||||
 QY 9374 AACGGGAGCTAACACTCCAGGCTTAAGCCATTTCTGTTTTTTTTT 9424
 Db |||||
 QY 9362 AACGGGAGCTAACACTCCAGGCTTAAGCCATTTCTGTTTTTTTTT 9412
 Db |||||

RESULT 10
 ABK91411
 ID ABK91411 standard; DNA; 9605 BP.
 XX AC ABK91411;
 XX AC
 DT 15-NOV-2002 (first entry)
 XX AC
 DE Hepatitis C virus Con 1 isolate DNA.
 XX AC
 KW HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
 KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
 KW internal ribosome entry site; IRES; NS5A; NS5B; HCV replication.
 XX AC
 OS Hepatitis C virus.
 XX AC
 PH Key Location/Qualifiers
 FT CDS 342..9374
 FT /+tag= a
 FT /product= "HCV polyprotein"
 FT /note= "The polyprotein consists of the Core, E1, E2, P7,
 NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
 XX AC

PN WO200259321-A2.
 XX 01-AUG-2002.
 XX 16-JAN-2002; 2002WO-EP000526.
 XX 23-JAN-2001; 2001US-0263479P.
 XX (RICE-) IST RIERCHE BIOL MOLECOLARE ANGELETTI.
 PA De Francesco R, Migliaccio G, Paonessa G;
 PI WPI; 2002-599793/64.
 XX P-PSDB; ABG32451.
 PT New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
 PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
 PT ribosome entry site (IRES) region, useful in studying HCV replication and
 PT expression.
 XX Claim 9; Page 36-39; 69pp; English.
 PS The invention relates to nucleic acid molecules comprising altered HCV
 CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
 CC internal ribosome entry site (IRES) region coding for one or more NS3,
 CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
 CC are detailed in the specification. Also included are (1) an expression
 CC vector comprising a nucleotide sequence coding for the altered nucleic
 CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
 CC recombinant cell human hepatoma cell comprising the altered nucleic acids
 CC ; (3) a recombinant cell produced by introducing into a human hepatoma
 CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
 CC an HCV replicon enhanced cell or which containing a functional HCV replicon; (5)
 CC an HCV replicon enhanced cells made in the method; and (6) measuring the
 CC ability of a compound to affect HCV activity. The HCV replicons and HCV
 CC replicon enhanced cells are useful in studying HCV replication and
 CC expression, and HCV and host cell interactions, producing HCV RNA and
 CC proteins, and providing a system for measuring the ability of a compound
 CC to modulate one or more HCV activities e.g. to discover drugs which may
 CC treat HCV mediated diseases such as liver failure, cirrhosis and
 CC hepatocellular carcinoma. The present sequence is the HCV replicon Con 1,
 CC used as a basis for the adaptive mutations of the invention
 XX SQ Sequence 9605 BP; 1910 A; 2883 C; 2733 G; 2079 T; 0 U; 0 Other;
 Query Match 86.0%; Score 8250.2; DB 6; Length 9605;
 Best Local Similarity 91.3%; Pred. No. 0;
 Matches 8767; Conservative 0; Mismatches 828; Indels 10; Gaps 1;
 QY 1 GCCAGCCCCCTGATGGGGGCGACACTCCACCATGAATCACTCCCTCTGTGAGGAACACTG 60
 Db |||||
 QY 61 TCTTCACGAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTGTCGACGCTCCAGGAC 120
 Db |||||
 QY 61 TCTTCACGAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTGTCGACGCTCCAGGAC 120
 Db |||||
 QY 121 CCCCCCTCCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGAATTCGCGAG 180
 Db |||||
 QY 121 CCCCCCTCCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGAATTCGCGAG 180
 Db |||||
 QY 181 GACGACCGGCTCTCTTTTGGATCAACCGCTCAATGCTCGAGATTTGGGCGTGCCTCC 240
 Db |||||
 QY 181 GACGACCGGCTCTCTTTTGGATCAACCGCTCAATGCTCGAGATTTGGGCGTGCCTCC 240
 Db |||||
 QY 241 GCAGACTGCTAGCCGAGTAGTGTGGTCCGCGAAAGGCTTGTGTTACTGCTGATAGG 300
 Db |||||
 QY 241 GCAGACTGCTAGCCGAGTAGTGTGGTCCGCGAAAGGCTTGTGTTACTGCTGATAGG 300
 Db |||||
 QY 301 GTCTTTCGAGGTGCCCGGGGAGTCTCGTAGACCGGTGACCATGACGACGATCTTAAC 360
 Db |||||
 QY 301 GTCTTTCGAGGTGCCCGGGGAGTCTCGTAGACCGGTGACCATGACGACGATCTTAAC 360
 Db |||||

QY 361 CTCAGAGAAAACCAACGTAACACCAACCGCGCCCAACAGGACGTCAGATTCOCGGCG 420
Db 361 CTCAGAGAAAACCAACGTAACACCAACCGCGCCCAACAGGACGTCAGATTCOCGGCG 420
QY 421 GTGCTCAGATCGTGTGGAGTTTACCTGTTCGGCGCAGCGGCCACAGTTGGGTGTGC 480
Db 421 GTGCTCAGATCGTGTGGAGTTTACCTGTTCGGCGCAGCGGCCACAGTTGGGTGTGC 480
QY 481 GCGCGACTAGGAAGGCTTCGAGCGGTTCGCAACCTCGTGAAGCGCAGCAACCTATCCCAA 540
Db 481 GCGCGACTAGGAAGGCTTCGAGCGGTTCGCAACCTCGTGAAGCGCAGCAACCTATCCCAA 540
QY 541 AGGCTCGCCGACCCGAGGACAGGCTGGCTCAGCCCGGGTACCTTGGCCCTCTATG 600
Db 541 AGGCTCGCCGACCCGAGGCTGGCTCAGCCCGGGTACCTTGGCCCTCTATG 600
QY 601 GCAATGAGGCGCTGGGTGGGAGGATGCTCTGTGTCACCCCGGCTCCCGGCTAGTT 660
Db 601 GCAATGAGGCGCTGGGTGGGAGGATGCTCTGTGTCACCCCGGCTCCCGGCTAGTT 660
QY 661 GGGGCCCAACGACCCCGCGCTAGGTTCGTAACCTTGGGTAAAGTTCATCGATACCCCTTA 720
Db 661 GGGGCCCAACGACCCCGCGCTAGGTTCGTAACCTTGGGTAAAGTTCATCGATACCCCTTA 720
QY 721 CATGGGCTTCGCCATCTCATGGGTACATTCGGCTCGTCGGCGCCCGCTAGGGGGG 780
Db 721 CATGGGCTTCGCCATCTCATGGGTACATTCGGCTCGTCGGCGCCCGCTAGGGGGG 780
QY 781 CTGCCAGGCGCTTGGCACAGGCTGTCGGGTTCGGAGACGCGTGACTATGCAACAG 840
Db 781 CTGCCAGGCGCTTGGCACAGGCTGTCGGGTTCGGAGACGCGTGACTATGCAACAG 840
QY 841 GGAATCTGCCCGGTGCTCTTCTCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 900
Db 841 GGAATCTGCCCGGTGCTCTTCTCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 900
QY 901 TCCGAGCTTCGCTTATGAGTGCACAGCTGTCGGGTATACCATGTCAGACAGACT 960
Db 901 TCCGAGCTTCGCTTATGAGTGCACAGCTGTCGGGTATACCATGTCAGACAGACT 960
QY 961 GCTCAAATCAAGCAATGTGTATGAGGACGCGATGATCATGTATCTCCCGGTGCG 1020
Db 961 GCTCAAATCAAGCAATGTGTATGAGGACGCGATGATCATGTATCTCCCGGTGCG 1020
QY 1021 TGCCCTGTGTTGAGAGGTTAAGCTCCGCTGCTGGGTAGCGCTCATCTCCAGCTCG 1080
Db 1021 TGCCCTGTGTTGAGAGGTTAAGCTCCGCTGCTGGGTAGCGCTCATCTCCAGCTCG 1080
QY 1081 CGGCAGGAATGCCAGCGTCCCACTAGCAATAGGACGCGCACTGCTGCTGCTGCTG 1140
Db 1081 CGGCAGGAATGCCAGCGTCCCACTAGCAATAGGACGCGCACTGCTGCTGCTGCTG 1140
QY 1141 GGAAGGCTCTTCTGCTCGCTATGTATGAGGAGTCTCTGCGGATCTATTTCCCTCG 1200
Db 1141 GGAAGGCTCTTCTGCTCGCTATGTATGAGGAGTCTCTGCGGATCTATTTCCCTCG 1200
QY 1201 TCTCCAGCTGTACCTTCTGCTCGCTCGCGGCAAGACAGTGCAGACTGCACTGCT 1260
Db 1201 TCTCCAGCTGTGTACCTTCTGCTCGCTCGCGGCAAGACAGTGCAGACTGCACTGCT 1260
QY 1261 CAATCTATCCCGGCACTGATGAGTTCAGGTCACCGCTTGGATATGATGATGATGAT 1320
Db 1261 CAATCTATCCCGGCACTGATGAGTTCAGGTCACCGCTTGGATATGATGATGATGAT 1320
QY 1321 CACCTAACAGCCCTAGTGTGTCGAGTTGCTCCGGATCCCAAGCTGCTGTTGAGCA 1380
Db 1321 CACCTAACAGCCCTAGTGTGTCGAGTTGCTCCGGATCCCAAGCTGCTGTTGAGCA 1380
QY 1381 TGGTGGCGGGGCCCACTGGGGAGTCTGGGGGCTTGCCTACTATTCATGTTAGGGA 1440
Db 1381 TGGTGGCGGGGCCCACTGGGGAGTCTGGGGGCTTGCCTACTATTCATGTTAGGGA 1440
QY 1441 ACTGGGCTAAGGTTCTGATCTGGGCTACTCTTTGCGCGGCTTGAACGGGAGACCCACA 1500

Db 1441 ACTGGGCTAAGGTTCTGATGTGATGCTACTCTTTGCGCGGTTGACGGGGAACTATG 1500
QY 1501 CGACGGGAGGTTGGCGGCCACACACCTCCGGGTTCACGTCCTTTCTCATCTGGGG 1560
Db 1501 TGACAGGGGGGACGATGCGCAAAAACACCTCGGATTAACGTCCTCTTTTACCCGGGT 1560
QY 1561 CGTCTCAGAAATCAGCTTGTGAATACAAAGGAGCTGGACATCAACAGGATGCC 1620
Db 1561 CATCCAGAAATCAGCTTGTAAACACAAACGAGCTGGACATCAACAGGATGCC 1620
QY 1621 TAAATTCGAATGACTCCCTCCAAATGGGTTCCTTTGCGCGCTGTTTACGACACAAAGT 1680
Db 1621 TGAATTCGAATGACTCCCTCAACATGGGTTCCTTTGCTGCTGCTGCTGCTGCTGCTG 1680
QY 1681 TCAATTCGCTCCGGTTCGCGAGCGCATGGCGAGCTGGCGGCTGAGCGGCTTCCGCTC 1740
Db 1681 TCAATTCGCTCCGGTTCGCGAGCGCATGGCGAGCTGGCGGCTGAGCGGCTTCCGCTC 1740
QY 1741 AGGGTGGGGCCCATCACTATCACTAAGCTTAACAGCTCGSATCAGAGGCTTATTCGT 1800
Db 1741 AGGGTGGGGCCCATCACTTACATGATGATCAACAGCTCGSATCAGAGGCTTATTCGT 1800
QY 1801 GGCATTACCGGCTTCGACCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860
Db 1801 GGCATTACCGGCTTCGACCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860
QY 1861 ATTGTTTCAACCCCAAGCCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920
Db 1861 ACTGTTTCAACCCCAAGCCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920
QY 1921 ATAGTGGGGGAGAAATGAGACAGAGCTGATGCTCTCAACAAACAGCGCTCGCCACAG 1980
Db 1921 ACAGTGGGGGAGAAATGAGACAGAGCTGATGCTCTCTTAAACAAACAGCGCTCGCCACAG 1980
QY 1981 GCAACTGTGCTGGGTGTACATGGAATGATGATGATGATGATGATGATGATGATGATGAT 2040
Db 1981 GCAACTGTGCTGGGTGTACATGGAATGATGATGATGATGATGATGATGATGATGATGAT 2040
QY 2041 CCCCCTGTAAATCGGGGGGTTCGTAAACCGCATCTTGATCTGCCCCCAACGAGCTGCTTC 2100
Db 2041 CCCCCTGTAAATCGGGGGGTTCGTAAACCGCATCTTGATCTGCCCCCAACGAGCTGCTTC 2100
QY 2101 GGAAGCACCCGAGGCTTACACAAAATGCTGCTCGGGGCTGCTGCTGCTGCTGCTGCTGCT 2160
Db 2101 GGAAGCACCCGAGGCTTACACAAAATGCTGCTCGGGGCTGCTGCTGCTGCTGCTGCTGCT 2160
QY 2161 GCCTAGTAGACTACCCATACAGGCTTGGCACTACCCCTGCACTCTCAATTTTTCATCT 2220
Db 2161 GCCTAGTAGACTACCCATACAGGCTTGGCACTACCCCTGCACTCTCAATTTTTCATCT 2220
QY 2221 TTAAGGTTAGGATGATGCTGGGGGCTGGAGACAGGCTCAATGCGCATGCAATGGA 2280
Db 2221 TTAAGGTTAGGATGATGCTGGGGGCTGGAGACAGGCTCAATGCGCATGCAATGGA 2280
QY 2281 CTCGAGGAGCGCTGTAACTTGAGGACAGGATAGGTGAGACTCAGCCCTGCTGCTG 2340
Db 2281 CTCGAGGAGCGCTGTAACTTGAGGACAGGATAGGTGAGACTCAGCCCTGCTGCTG 2340
QY 2341 TGTCTAACAGAGTGCAGATCTGCCCTGTGTCTTTCACCAACCTACCCGCTTATCCA 2400
Db 2341 TGTCTAACAGAGTGCAGATCTGCCCTGTGTCTTTCACCAACCTACCCGCTTATCCA 2400
QY 2401 CTGCTTTGATCCATCTCCATCAGACATCGTGAGCTGCAATACCTGTACGTTAGGTT 2460
Db 2401 CTGCTTTGATCCATCTCCATCAGACATCGTGAGCTGCAATACCTGTACGTTAGGTT 2460
QY 2461 CAGGTTTGTCTCTCTTTGCAATCAATGGGAGTACATCTGTTGCTTTTCTCTCTCTG 2520
Db 2461 CAGGTTTGTCTCTCTTTGCAATCAATGGGAGTACATCTGTTGCTTTTCTCTCTCTG 2520
QY 2521 CAGACCGCGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580
Db 2521 CAGACCGCGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580

Db 2521 CGGACCGCGCGTCTGTGCTGCTGTGTGATGATGCTGCTGATAGCTCAAGCTGAGCGG 2580
Qy CTTTAGAGAACTTGGTGTCTCAATGCGGGTCCGCTGCGCGAGCGATGATTTCT 2640
Db CTTTAGAGAACTTGGTGTCTCAATGCGGGTCCGCTGCGCGAGCGATGATTTCT 2640
Qy CTTTCTGTGTCTTCTGTGCGCGCTGTGATCAATAAGGGGAGCTGCTTCTGGGCGG 2700
Db CTTTCTGTGTCTTCTGTGCGCGCTGTGATCAATAAGGGGAGCTGCTTCTGGGCGG 2700
Qy CGTATGCTTTTANGGCTATGGCGGTGCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 2760
Db CATATGCT 2760
Qy CTTAGCGCTTGGACCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820
Db CATACGCTTGGACCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820
Qy TATCTTGAACCTTGTACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
Db TATCTTGAACCTTGTACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
Qy AATATCTTATACACGAGCGGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940
Db AATATCTTATACACGAGCGGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940
Qy GGGAGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
Db GGGAGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
Qy ACATCAACCAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3060
Db ACATCAACCAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3060
Qy CGAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3120
Db CGAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3120
Qy AAGTGGCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
Db AAGTGGCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
Qy CPTACGTTTATACCACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3240
Db CPTACGTTTATACCACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3240
Qy TTGCGGTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3300
Db TTGCGGTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3300
Qy GAGCAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3360
Db GAGCAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3360
Qy GGAAGGAGATATTTTGGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3420
Db GGAAGGAGATATTTTGGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3420
Qy CGCCCATCGCGCTACTCCCAACAAACGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 3480
Db CGCCCATCGCGCTACTCCCAACAAACGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 3480
Qy TCACAGCGCGGACAGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3540
Db TCACAGCGCGGACAGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3540
Qy AATCTTCTCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3600
Db AATCTTCTCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3600
Qy CGAAGACCTTAGCGGCTCAAAAGGTCCTCAATCAACCAATGTACCAATGTAGACCTGG 3660
Db CGAAGACCTTAGCGGCTCAAAAGGTCCTCAATCAACCAATGTACCAATGTAGACCTGG 3660

Qy 3661 ACCTCTGCGCTGGCAGGCGCGCGCGCGCGCGCGCTGCCATGACACCATGAGCTGTGGCA 3720
Db 3661 ACCTCTGCGCTGGCAGGCGCGCGCGCGCGCGCGCTGCCATGACACCATGAGCTGTGGCA 3720
Qy 3721 GCTCGGACCTTTACTTTGGTCAAGACATGCTCATGTCTATCCGCTGCGCGCGCGGAGCG 3780
Db 3721 GCTCGGACCTTTACTTTGGTCAAGACATGCTCATGTCTATCCGCTGCGCGCGGAGCG 3780
Qy 3781 ACAGCAGGAGTGTACTTCTCTCCCAAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3840
Db 3781 ACAGCAGGAGTGTACTTCTCTCCCAAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3840
Qy 3841 GTCCATTTCTTCTCTCGGGGCACTGCTGCTGCGGCTCTCTCTCTCTCTCTCTCTCTCT 3900
Db 3841 GTCCATTTCTTCTCTCGGGGCACTGCTGCTGCGGCTCTCTCTCTCTCTCTCTCTCT 3900
Qy 3901 GGGGCTCTCGGAGGCGGTGGACTTCTATCCGCTTGGTCTATGGAATACCATCGGT 3960
Db 3901 GAGGGCTCTCGAAGGCGGTGGACTTCTATCCGCTTGGTCTATGGAATACCATCGGT 3960
Qy 3961 CTCCGCTCTCTCAGACAACTCAACCCCGCGCTGTACCGGAGCATTTCCAAAGTGGCAC 4020
Db 3961 CTCCGCTCTCTCAGACAACTCAACCCCGCGCTGTACCGGAGCATTTCCAAAGTGGCAC 4020
Qy 4021 ATCTGACACCTCTCTCTGCGCAGCGGCAAGACCAAAAGTGGCGGCTGTATGACGCG 4080
Db 4021 ATCTGACACCTCTCTCTGCGCAGCGGCAAGACCAAAAGTGGCGGCTGTATGACGCG 4080
Qy 4081 AAGGTATTAAGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4140
Db 4081 AAGGTATTAAGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4140
Qy 4141 ATATGTCGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4200
Db 4141 ATATGTCGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4200
Qy 4201 CGGGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4260
Db 4201 CGGGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4260
Qy 4261 GGGGCGCTTATGACATCATATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Db 4261 GGGGCGCTTATGACATCATATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Qy 4321 TGGGCTATCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4380
Db 4321 TGGGCTATCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4380
Qy 4381 CCACCGCTACCTCGGGATCGGTTACCGTCCACACCCCAATATCGAGGAAATAGGCC 4440
Db 4381 CCACCGCTACCGTCCGGATCGGTTACCGTCCACACCCCAATATCGAGGAAATAGGCC 4440
Qy 4441 TGTCCCAATAGGAGATCCCTCTCTATGGAAGGCAATCCCAATGAGGCAATCAGG 4500
Db 4441 TGTCCCAATAGGAGATCCCTCTCTATGGAAGGCAATCCCAATGAGGCAATCAGG 4500
Qy 4501 GGGGAGGCTATCTCATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4560
Db 4501 GGGGAGGCTATCTCATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4560
Qy 4561 TGACAGGCTCGGACTGAACTGATGATGATGATGATGATGATGATGATGATGATGATG 4620
Db 4561 TGACAGGCTCGGACTGAACTGATGATGATGATGATGATGATGATGATGATGATGATG 4620
Qy 4621 GCGCTATCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4680
Db 4621 GCGCTATCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4680
Qy 4681 ATTTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4740
Db 4681 ATTTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4740

Db	6901	AGCTAGTCTGGCAGGGGATCTCCCCCTCTCTTGGCCAGCTCATAGCTAGCAGCTGT	6961
Qy	6961	CTGCGCTTCTTTGAAGCGCATGCACTACCCACCATGATCTCCCGGACGCTGACCTCA	7020
Db	6961	CTGCGCTTCTCTTGAAGCAATATGCACTACCGTCTATGACTCTCCCGGACGCTGACCTCA	7020
Qy	7021	TCGAGGCCAACCTCTTGTGGCGGACGAGATGSGCGGAACATCATCTCGTGTGGAGTCAG	7080
Db	7021	TCGAGGCCAACCTCTGTGGCGGACGAGATGSGCGGAACATCATCCCGGTGTGGAGTCAG	7080
Qy	7081	AGAAATAAGGTAGTAATTTGTGACTCTTTTCGAACCGCTTTCACGCGGAGGGGATGAGAGGG	7140
Db	7081	AAAAATAAGGTAGTAATTTTGGACTCTTTTCGAGCCGCTCCCAAGCGGAGGAGATGAGAGGG	7140
Qy	7141	AGATATCGTTCGGCGGCGAGATCTCTGGAAAAATCCAGGAATTCCTCTAGCGTTGCCCA	7200
Db	7141	AAGTATCCGTTTCCGCGGAGATCTGTGGAGGTCCAGAAAATTCCTCTGAGCGATGCCCA	7200
Qy	7201	TATGGGCACGCCGGAATACAAATCCTCCACTGCTAGAGTCTGTGAAGAGACCGGACTACG	7260
Db	7201	TATGGGCACGCCGGAATTACACCTCCACTGTTAGAGTCTGTGAAGAGACCGGACTACG	7260
Qy	7261	TCCCTCCGGTGGTACACGATGCCCATTTGCCACCTACCAAGGTCTCTCAATACCACTC	7320
Db	7261	TCCCTCCAGTGGTACACGGGTGTCCATTGCGCGCTTGCACAGGCCCTCCGATACCACTC	7320
Qy	7321	CACGGAAAGAGACGGTTCCTCTGACAGAAATCCAATGTGTCTTTCGCTTGGCGGAGC	7380
Db	7321	CACGGAGAGAGACGGTTCCTCTGCAGAAATCTACGTTGTCTTTCGCTTGGCGGAGC	7380
Qy	7381	TCGCCACTAAGACCTTCGGTAGTTCGGATGTCGGCGTGTATAGCGGACAGCGGACCG	7440
Db	7381	TCGCCCAAAACCTTCGGCAGTCCGAAATCGTCGGCGCTCGACAGCGGACAGCGCAACGG	7440
Qy	7441	CCCTTCTGTACTGCTCTCGACACGGTGACAAAGGATCCGACGTTTGTAGTCTCTCT	7500
Db	7441	CCCTCTCTGACCAAGCTTCGACACGGGACGCGGATCCGACGTTTGTAGTCTCTCTCT	7500
Qy	7501	CCATGCCCCCTTTGAAGGGAGCCGGGGACCCCGATCTCAGCGACGGGTCTTGGTCTA	7560
Db	7501	CCATGCCCCCTTTGAGGGGAGCCGGGGATCCCGATCTCAGCGACGGGTCTTGGTCTA	7560
Qy	7561	CCGTGAGTGAGGAGCTAGTGAGATGTGTCTCTCTCTCAATGTCTCTATGCTGTGCACAG	7620
Db	7561	CCGTAGCGAGGAGCTAGTGAGACGTGTCTCTCTGTGATGTCTACACATGTGCACAG	7620
Qy	7621	GCGCCCTGATCACGCCATGCGCTCGGAGGAAGATTAAGCTGCCCATCAACCCGTTGAGCA	7680
Db	7621	GCGCCCTGATCACGCCATGCGCTCGGAGGAACCAAGCTGCCCATCAATGCTACTGAGCA	7680
Qy	7681	ACTCTTTTCTGCTCACCACAACTTGTCTTAGCGCAAAATCCCGCAGGCGCAAGCTCC	7740
Db	7681	ACTCTTTTCTGCTCACCACAACTTGTCTTAGCTATGTCTACAACTCTCGCAGGCGCAAGCTCC	7740
Qy	7741	GGCAGAGAAGGTCACTTTGACAGATTGCAAGTTCCTGGATGATCACTACGGGACGTAC	7800
Db	7741	GGCAGAGAAGGTCACTTTGACAGACTGAGAGTCTTGAGCAGCACTACGGGACGTGC	7800
Qy	7801	TCAAGAGATGAAGGCGAGCGTCCACAGTTAAGGCTTAAGCTTCTATCTATAGAGGAGG	7860
Db	7801	TCAAGGAGATGAAGGCGAAGCGTCCACAGTTAAGGCTTAAGCTTCTATCGTGGAGGAG	7860
Qy	7861	CCTCAAGCTGACGCCCCCACTTCGGCCAAATCCAAATTTGGCTATGSGGCAAGGACG	7920
Db	7861	CCTGTAGCTGACGCCCCCACTTCGGCCAGATCTAAATTTGGCTATGSGGCAAGGACG	7920
Qy	7921	TCCGGAACTTACGACGAGCGGCTTAACCACTCCGCTCCGTTGTGGAGGACTTGTCTGG	7980
Db	7921	TCCGGAACTTACGACGAGCGGCTTAACCACTCCGCTCCGTTGTGGAGGACTTGTCTGG	7980
Qy	7981	AAGACACTGAACCAATTTGACACACCATCATGGCAAAAGTAGAGTTTCTGTGGTCC	8040
Db	7981	AAGACACTGACCAATTTGACACCACTCATGGCAAAATTTGAGTTTCTGTGGTCC	8040

Db 481 GCGCGACTAGAGAGACTTCCGAGCGGTCGCAACCTCGTGAAGGCGCAACACTATCCCA 540
QY 541 AGGCTCCGACCCGAGGCGAGGCTTGGGCTAGCCCGGAGTACCTTGGCCCTCTATG 600
Db 541 AGGCTCCGACCCGAGGCGAGGCTTGGGCTAGCCCGGAGTACCTTGGCCCTCTATG 600
QY 601 GCAATGAGGCGCTTGGGCTGGGCGAGGCTGGCTCCGTCACCCCGGCGCTCCGCGCTAGTT 660
Db 601 GCAATGAGGCGCTTGGGCTGGGCGAGGCTGGCTCCGTCACCCCGGCGCTCCGCGCTAGTT 660
QY 661 GGGGCCCAAGGACCCCGGCGTAGTGGCTTAATTTGGGTAAAGTCAATCGATCCCTTA 720
Db 661 GGGGCCCAAGGACCCCGGCGTAGTGGCTTAATTTGGGTAAAGTCAATCGATCCCTTA 720
QY 721 CATGCGGCTTCCGCGATCTCATGGGTACATTCGCTCGTCCGCGCCCTTAGGGGGCG 780
Db 721 CGTGGCGCTTCCGCGATCTCATGGGTACATTCGCTCGTCCGCGCCCTTAGGGGGCG 780
QY 781 CTGCCAGGCGCTTGGCACACAGGCTCCGCGTTCTGGAGGACGGCTCAACTATGCAACAG 840
Db 781 CTGCCAGGCGCTTGGCACATGGCTCCGCGTTCTGGAGGACGGCTCAACTATGCAACAG 840
QY 841 GGAATTTGCCGGTGTCTTTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db 841 GGAATTTGCCGGTGTCTTTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
QY 901 TCCAGCTTCCGCTTATGAAGTGGCGAAGTGTCCGGGATACCACTGTCAGCAAGCACT 960
Db 901 TCCAGCTTCCGCTTATGAAGTGGCGAAGTGTCCGGGATACCACTGTCAGCAAGCACT 960
QY 961 GCTCCAACTCAAGCAATTTCTATGAGGACAGGAGCTGATCATGATCACTCCCGGCTGG 1020
Db 961 GCTCCAAAGCAAGCTTGTATGAGGACAGGAGCTGATCATGATCACTCCCGGCTGG 1020
QY 1021 TGGCTGTGTCAGAGGTAACAGCTCCGCTTCTGCTGAGTACGCTCACTCCCAAGCTCG 1080
Db 1021 TGGCTGTGTCAGAGGTAACAGCTCCGCTTCTGCTGAGTACGCTCACTCCCAAGCTCG 1080
QY 1081 CGGCGAGGATGCCAGGTCGCCACTAGCAATACGAGCGCACTGCTGCTGCTGCTG 1140
Db 1081 CGGCGAGGATGCCAGGTCGCCACTAGCAATACGAGCGCACTGCTGCTGCTGCTG 1140
QY 1141 GGAAGCTGCTTCTGCTCCGCTATGACGTTGGGATCTCTGCGGATCTATTTCTCTCG 1200
Db 1141 GGAAGCTGCTTCTGCTCCGCTATGACGTTGGGATCTCTGCGGATCTATTTCTCTCG 1200
QY 1201 TCTCCAGCTGTTCACTTCTCGCTCCGCGATGAGACAGTGCAGGACTGCACTGCT 1260
Db 1201 TCTCCAGCTGTTCACTTCTCGCTCCGCGATGAGACAGTGCAGGACTGCACTGCT 1260
QY 1261 CAATCTATCCCGGCAATGATCAGTCACTGCTCCGCTTGGGATGATGATGAACTGGT 1320
Db 1261 CAATATATCCCGGCAATGATCAGTCACTGCTCCGCTTGGGATGATGATGAACTGGT 1320
QY 1321 CACTATACAGCCCTAGTGTGTCAGTTGCTCCGATGCCAAGCTGCTGTGAGCA 1380
Db 1321 CACTATACAGCCCTAGTGTGTCAGTTGCTCCGATGCCAAGCTGCTGTGAGCA 1380
QY 1381 TGGTGG 1440
Db 1381 TGGTGG 1440
QY 1441 ACTGGGCTAAGGTTCTGATTTGGGCTACTCTTTTGGCGGGTGGAGGGGAGAGCCACA 1500
Db 1441 ACTGGGCTAAGGTTCTGATTTGGGCTACTCTTTTGGCGGGTGGAGGGGAGAGCCATG 1500
QY 1501 CGACGGGAGGGTGGCGGGGCAACACCTCGGGTCAAGCTCCCTTTCTCTCTCTCTCT 1560
Db 1501 TGAAGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
QY 1561 CGTCTCAGAAATCAGCTTGTGATACCAAGCGGAGGCTGCGCATCAAGAGCTGCC 1620

Db 1561 CATCCAGAGAAATCCAGCTTGTAAACCAACGAGGAGCTGGCACATCAACAGAGCTGCC 1620
QY 1621 TAAATTTGAAATGACTCCCTCCAAACTGGGTTCTTTGCGCGCTGTTTATGACACAAAGT 1680
Db 1621 TGAATTTGAAATGACTCCCTCCAAACTGGGTTCTTTGCGCGCTGTTTATGAGTGCAGAGT 1680
QY 1681 TCAACTCTGTCGCGGTCGCGGAGCGCATGCCAGCTGCGAGCCCATCGAGCGCTTCGTC 1740
Db 1681 TCAACTCATCTGATGCCAGAGCGCATGCCAGCTGCGAGCCCATCGAGCGCTTCGTC 1740
QY 1741 AGGGTGG 1800
Db 1741 AGGGTGG 1800
QY 1801 GGCATTTACCGGCTCGACCGTGTGGTGTCTGTCGCGCTCGACAGGTTGTGTGTCAAGTGT 1860
Db 1801 GGCATTTACCGGCTCGACCGTGTGGTGTCTGTCGCGCTCGACAGGTTGTGTGTCAAGTGT 1860
QY 1861 ATTGTTTCAACCCCAAGCCCTGTTGTGGTGGGACCAACCGATGTTTCCGGTGTCCCTACGT 1920
Db 1861 ACTGCTTCAACCCCAAGCCCTGTTGTGGTGGGACCAACCGATGTTTCCGGTGTCCCTACGT 1920
QY 1921 ATAGCTGGGGGAGATGACAGAGCTGATGCTCTCTCAACAGCGCTTCGCCACAG 1980
Db 1921 ACAGTTGGGGGAGATGACAGAGCTGATGCTCTCTCAACAGCGCTTCGCCACAG 1980
QY 1981 GCAACTGGTTCGGCTGTACATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
Db 1981 GCAACTGGTTCGGCTGTACATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
QY 2041 CCCCCTGTAACTCGGGGGGTCGGTAAACCGCACTTGTATGTCGCCCAAGGAGCTGCTTC 2100
Db 2041 CCCCCTGTAACTCGGGGGGATCGGCAATAAACCCTTGACCTGCCCCCAAGGAGCTGCTTC 2100
QY 2101 GGAAGCACCCGAGGCTACTTACACAAATGTTGGCTCGGGGGCTGTTGTTGACACCTAGT 2160
Db 2101 GGAAGCACCCGAGGCTACTTACACAAATGTTGGCTCGGGGGCTGTTGTTGACACCTAGT 2160
QY 2161 GCCTAGTACTACCCATACAGGCTTGGCACTACCCCTGCACTCTCAATTTTTCATCT 2220
Db 2161 GCCTAGTACTACCCATACAGGCTTGGCACTACCCCTGCACTCTCAATTTTTCATCT 2220
QY 2221 TTAAGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280
Db 2221 TTAAGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280
QY 2281 CTCGAGGAGCGCTGTTAACTTGGAGGACAGGATGATGATGATGATGATGATGATGATGATG 2340
Db 2281 CTCGAGGAGCGCTGTTAACTTGGAGGACAGGATGATGATGATGATGATGATGATGATGATG 2340
QY 2341 TGTCTACACAGAGTGGCAGATCTGCTGCTGCTTTCACACCCCTTACCGGCTTTATCCA 2400
Db 2341 TGTCTACACAGAGTGGCAGATCTGCTGCTGCTTTCACACCCCTTACCGGCTTTATCCA 2400
QY 2401 CTGGTTTGAATCCATCTCCATCAGAACTGCTGGAAGCTGATGATGATGATGATGATGATG 2460
Db 2401 CTGGTTTGAATCCATCTCCATCAGAACTGCTGGAAGCTGATGATGATGATGATGATGATG 2460
QY 2461 CAGGCTTGTCTCTCTTGAATCAATGAGGATGATGATGATGATGATGATGATGATGATGATG 2520
Db 2461 CAGGCTTGTCTCTCTTGAATCAATGAGGATGATGATGATGATGATGATGATGATGATGATG 2520
QY 2521 CAGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
Db 2521 CAGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
QY 2581 CCTTAGAGAACTTGGTGGTCTCAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640
Db 2581 CCTTAGAGAACTTGGTGGTCTCAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640
QY 2641 CTTTTCTGTTGTTCTTCTGCGCGCTGTTTACATTAAGGCGAGGCTGCTGCTGCTGCTGCTG 2700
Db 2641 CTTTTCTGTTGTTCTTCTGCGCGCTGTTTACATTAAGGCGAGGCTGCTGCTGCTGCTGCTG 2700

QY 2701 CGTATGCTTTTATGGCGTATGGCGCTGCTCCTGCTCCTACTGGGGTTTACACACGAG 2760
DB |||||
QY 2701 CATATGCCCTTACGGCGTATGGCGCTACTCTGCTCTGCTGGGTTTACACACGAG 2760
DB |||||
QY 2761 CTTAGCCCTTGACCGGGAGATGGTGATCTGCTGGGGGGTGGGTTCTGTAGCTCTGG 2820
DB |||||
QY 2761 CATACGCCATGACCGGGAGATGGGAGCATCTGGGAGGCGGGTTTTCGTAGGCTGA 2820
DB |||||
QY 2821 TATCTCTGACCTTGTCAACCATACTCAAAAGTGTCTCTCACTAGGCTCATATGCTGGTTAC 2880
DB |||||
QY 2821 TACTCTTGACCTTGTACCGGCATATAGCTGTTCTCTAGGCTCATATGCTGGTTAC 2880
DB |||||
QY 2881 AATATCTTATACACAGCGAGGGGACATCAAGATGTGGTTCGCCGCCCTTCAAGTTTC 2940
DB |||||
QY 2881 AATATTTTATCACAGGGCCGAGGACACTTTCGAAGTGTGGATCCGCCCTTCAAGTTTC 2940
DB |||||
QY 2941 GGGGAGCGCGATGCCATCATCTCTCTCTCACTGCTGCGGTTTATCAGAGTTAATTTTG 3000
DB |||||
QY 2941 GGGGGGCGCGATGCCGTCATCTCTCTCACTGCTGCGGATCCACGAGCTTAATCTTTA 3000
DB |||||
QY 3001 ACATCACCAAACTCTGCTGCGCATACTCGGCCGCTCATGTGTCTCCAGGCTGGCATAA 3060
DB |||||
QY 3001 CCATCACCAAACTCTGCTGCGCATACTCGGTCCTCATGTGTCTCCAGGCTGGTATTA 3060
DB |||||
QY 3061 CGAGAGTCCGCTACTTCTGCTGCGGCTCAAGGCTCATCTGCTCATGCTATGTAGTCCGAA 3120
DB |||||
QY 3061 CCAAGTCCGCTACTTCTGCTGCGGACACCGGCTCATCTGCTCATGCTATGTAGTCCGAA 3120
DB |||||
QY 3121 AAGTCGCGGGGCTCATATGTGCTCAAAATGCTTCTCATGAGCTGGCGGCTGACAGGTA 3180
DB |||||
QY 3121 AGGTTGCTGGGGTCAATATGTCCTCAAAATGCTTCTCATGAGTGGCGGCTGACAGGTA 3180
DB |||||
QY 3181 CTTAGCTTTATTAACCATCTTACCCACTGCGGCACTGGGCCACCGGCTTACGAGCC 3240
DB |||||
QY 3181 CTTAGCTTTATTAACCATCTTACCCACTGCGGCACTGGGCCACCGGCTTACGAGCC 3240
DB |||||
QY 3241 TTGCGGTGGCGTAGAGCCCGTCTCTCTCGGCACTGGGCACTGAGGCTTCACTCTGG 3300
DB |||||
QY 3241 TTGCGGTGGCGTAGAGCCCGTCTCTCTCGGCACTGGGCACTGAGGCTTCACTCTGG 3300
DB |||||
QY 3301 GAGCAGACACCGCTGCTGGGCACTCATCTTGGGCTTACCGCTCTCGGCCCGAGGG 3360
DB |||||
QY 3301 GGGCAGACACCGCGCTGCTGGGCACTCATCTTGGGCTTACCGCTCTCGGCCCGAGGG 3360
DB |||||
QY 3361 GGAAGGAGATATTTTGGGACCGGCTGATGATGCTCGAAGGGCAAGGGTGGGCTCTCTG 3420
DB |||||
QY 3361 GGAGGAGATACATCTGGGACCGGCAAGAGCTTGAAGGGCAGGGTGGGCTCTCTG 3420
DB |||||
QY 3421 CGCCCATACCGGCTTCTCCCAACAAACGGGGGCTTCTGCTTGGTGGATCACTAGCC 3480
DB |||||
QY 3421 CGCCTATTACGGCTTCTCCCAACAAACGGGGGCTTCTGCTTGGTGGATCACTAGCC 3480
DB |||||
QY 3481 TCACAGGCGGACAAAGAACAGGTCGAGGGGAGGTTCAAGTGGTTCTACCGCAACAC 3540
DB |||||
QY 3481 TCACAGGCGGACAAAGAACAGGTCGAGGGGAGGTTCAAGTGGTTCTACCGCAACAC 3540
DB |||||
QY 3541 AATCTTTCTGCGGACCTTCTCATCAAGCGGCTGCTGGAAGTCTTACCATGGCGCTGGCT 3600
DB |||||
QY 3541 AATCTTTCTGCGGACCTTCTCATCAAGCGGCTGCTGGAAGTCTTACCATGGCGCTGGCT 3600
DB |||||
QY 3601 CGAAGACCTTAGCCGGTTCAAAGTTCATCAACCAATGTACCAATGTAGACCTTGG 3660
DB |||||
QY 3601 CAAGAACCTTTCGCCGCCCAAGGGCCCAATCAACCAATGTACCAATGTAGACCTTGG 3660
DB |||||
QY 3661 ACTCGTGGCTGGCAGGCGCCCGGGGCGGCTCCATGACACATGCACTGTGCA 3720
DB |||||
QY 3661 ACTCGTGGCTGGCAGGCGCCCGGGGCGGCTTCTTGAACCATGCACTGTGCA 3720
DB |||||
QY 3721 GCTCGGACCTTTACTTGGTTCAGAGACATGCTGATGTCATCTCGGCTGGCGGCGAGGCG 3780
DB |||||
QY 3721 GCTCGGACCTTTACTTGGTTCAGAGGATGCGGATGTCATCTCGGCTGGCGGCGGCGG 3780
DB |||||

QY 3781 ACAGCAGGCGAAGTCTACTCTCCCGCAGCGCCGCTCCTACTGAAAGGCTCCTCGGGTG 3840
DB |||||
QY 3781 ACAGCAGGCGAAGTCTACTCTCCCGCAGCGCCGCTCCTACTGAAAGGCTCCTCGGGTG 3840
DB |||||
QY 3841 GTCCATTCCTTTGCGCTTGGGGGACGCTGCTGGGGCTCTTCCGGGCTGCTGTGTCACCC 3900
DB |||||
QY 3841 GTCCATTCCTTTGCGCTTGGGGGACGCTGCTGGGGCTCTTCCGGGCTGCTGTGTCACCC 3900
DB |||||
QY 3901 GGGGGGTCGGAAGCGGTGACATTCATACCGTGTGAGTCTATGGAATCACTACCTGCGGT 3960
DB |||||
QY 3901 GAGGGGTTTCCGAAGCGGTGACATTTGACCGCTGAGTCTATGGAATCACTATGCGGT 3960
DB |||||
QY 3961 CTCGGGCTTTCACAGACAATCAACCCCCCGGGTGTACCGCAGACATTTCCAAAGTGGCAC 4020
DB |||||
QY 3961 CCGCGGCTTTCAGGACAATCGTCCCTCCGGCGGTACCGCAGACATTTCCAGGTGGGCC 4020
DB |||||
QY 4021 ATCTGCAGCTTCTACTTGGCAGCGGCAAGAGACCAAAAGTCCGGCTGCGTATGAGGCC 4080
DB |||||
QY 4021 ATCTGCAGCTTCTACTTGGCAGCGGCAAGAGACCAAAAGTCCGGCTGCGTATGAGGCC 4080
DB |||||
QY 4081 AAGGTTACAAGTGTCTCTCTGTAACCGGTCCTGTCGCCGCCACCTTAGGGTTTGGGGGCT 4140
DB |||||
QY 4081 AAGGTTACAAGTGTCTCTCTGTAACCGGTCCTGTCGCCGCCACCTTAGGGTTTGGGGGCT 4140
DB |||||
QY 4141 ATATGTCCTAAGGCAATGATGACCCCTAATACAGAACCGGGTAAAGGACCATCAACA 4200
DB |||||
QY 4141 ATATGTCCTAAGGCAATGATGACCCCTAATACAGAACCGGGTAAAGGACCATCAACA 4200
DB |||||
QY 4201 CGGGGGCTTCCATTTACGTAATTCACCTTATGGAAGTTCCTTGGCGAGGTTGGTTTGG 4260
DB |||||
QY 4201 CGGGGGCTTCCATTTACGTAATTCACCTTATGGAAGTTCCTTGGCGAGGTTGGTTTGG 4260
DB |||||
QY 4261 GGGGGCTTATGACATCATATAATATGATGATGTCCTCACTCACTGACCTACCACTCT 4320
DB |||||
QY 4261 GGGGGCTTATGACATCATATAATATGATGATGTCCTCACTCACTGACCTACCACTCT 4320
DB |||||
QY 4321 TGGGCATCGGCACAGTCTCTGGAACCAAGCGGAGACGGCTGGAGCGGCTCTGCTGCTG 4380
DB |||||
QY 4321 TGGGCATCGGCACAGTCTCTGGAACCAAGCGGAGACGGCTGGAGCGGCTCTGCTGCTG 4380
DB |||||
QY 4381 CCAAGGCTTACCTTCCGGGATCGGTTACGTCGCAACCCCAATATCGAGGAATAGGCC 4440
DB |||||
QY 4381 CCAAGGCTTACCTTCCGGGATCGGTTACGTCGCAACCCCAATATCGAGGAATAGGCC 4440
DB |||||
QY 4441 TGTCCAAACATGAGAGAGATCCCTTCTATGCAAGGACATCCCATTTGAGGCGCTTCAAGG 4500
DB |||||
QY 4441 TGTCCAAACATGAGAGAGATCCCTTCTATGCAAGGACATCCCATTTGAGGCGCTTCAAGG 4500
DB |||||
QY 4501 GGGGGAGGCACTCTCATTTTCTGCCATTCCAAGAAAGAAATGTGACGAGCTCCGCGCAAGC 4560
DB |||||
QY 4501 GGGGGAGGCACTCTCATTTTCTGCCATTCCAAGAAAGAAATGTGATGAGCTCCGCGCAAGC 4560
DB |||||
QY 4561 TGACAGGCTCGGAGTCACTGACCTGCTAGCATATTAACGGGGCTTGTGATGCTCCGTCATAC 4620
DB |||||
QY 4561 TGTCCGGCTCGGAGTCACTGACCTGCTAGCATATTAACGGGGCTTGTGATGCTCCGTCATAC 4620
DB |||||
QY 4621 CGCCTATCCGAGAGACGCTGCTGCTGTCGCAACAGACGCTCTAATGACGGGTTTCAACCGCG 4680
DB |||||
QY 4621 CAACCTAGCGGAGACGTCATTTGCTGCTAGACGAGCGCTCTAATGACGGGCTTACCGCG 4680
DB |||||
QY 4681 ATTTTGAATCAAGTCACTGCAATACATGTGTCAACCCAGACAGTCACTTCAAGTCTGG 4740
DB |||||
QY 4681 ATTTTGAATCAAGTCACTGCAATACATGTGTCAACCCAGACAGTCACTTCAAGTCTGG 4740
DB |||||
QY 4741 ATCCCACTTCACTTGAACGAGCAGCGCTGCCCGCAGAGCGGGTTCGGGCTCGCAAC 4800
DB |||||
QY 4741 ACCCGACTTCACTTGAACGAGCAGCGCTGCCCGCAGAGCGGGTTCAGCTTCGAGC 4800
DB |||||
QY 4801 GCGAGGTAGAACTTGGCAGGGGTAGGAGTGGCATCTTACAGGTTTGTGACTTCCAGGAGAC 4860
DB |||||
QY 4801 GCGAGGTAGAACTTGGCAGGGGTAGGAGTGGCATCTTACAGGTTTGTGACTTCCAGGAGAC 4860
DB |||||
QY 4861 GGGGCTCGGGCAGTTCGATTTCTTCTGCTGCTGTGAGTGTGATGACGCGGGGCTGTGCTT 4920
DB |||||

4861 GGCCTCGGCATGTTTCAATTCCTCGGTTCTGTGCGAGTCTATGACGGGGCTGTGCTT 4920
4921 GGTATAGGCTCAGCGCGCTGAGACCTCGGTAGGTTCGGGCTTACCTAAATACACCCAG 4980
4921 GGTACAGCTCAGCGCGCGGAGACCTCAGTTAGGTTCGGGCTTACCTAAACACACCCAG 4980
4981 GGTTCGCCGTCTGCCAGGACCATCTGGAGTTCGGGAGAGCGTCTTTACAGGCTCAGCC 5040
4981 GGTTCGCCGTCTGCCAGGACCATCTGGAGTTCGGGAGAGCGTCTTTACAGGCTCAGCC 5040
5041 ACATAGATGCCACTTCCTGCTCCAGACTAAACAGGAGAGAGCAACTTTCCTTACTGG 5100
5041 ACATAGAGCGCCATTTCTTCTCCAGACTAAGCAGGAGAGCAACTTTCCTTACTGG 5100
5101 TGGCATATCAAGCTACAGTTCGGCCAGGCTCAAGCTCCACCTCCATCGTGGAGCCAAA 5160
5101 TAGCATACAGGCTACGTTGCGCCAGGCTCAGGCTCCACCTCCATCGTGGAGCCAAA 5160
5161 TGTGGAAGTCTCATACGGCTGAAACCTACACTGCAAGGCGCCAAACGCTCTGTGTATA 5220
5221 GGTAGAGCGGCTCCAAAATGAGGTCTATCTCAGACACCCCAATAACTAATACATG 5280
5221 GGTGGAGCGCTTCAAAACAGGTTACTACCAACACACACCCCAATAACTAATACATG 5280
5281 CATGCATCGGCTGACCTGAGGTCTGCTACTAGCACCTGGTCTGGTAGGGGAGTCC 5340
5281 CATTGATCGGCTGACCTGAGGTCTGCTACTAGCACCTGGTCTGGTAGGGGAGTCC 5340
5341 TTGCAGCTTTGGCGCATATGCTGCTGACAGAGCAGTGTGGTCAATTTGGGAGGATCA 5400
5341 TAGCAGCTCTGGCGCGCTATTGCTGCTGACAAACAGGAGCGTGGTCAATTTGGGAGGATCA 5400
5401 TCTTGTCCGGAGACGAGTGTGTTCCGACAGGAGGAGTCTCTACAGAGGTTGATG 5460
5401 TCTTGTCCGGAGACGCGCATATTTCCGACAGGAGGAGTCTCTTACGGGAGTTCGATG 5460
5461 AGATGGAAGAGTGTGCTTCAAACTCTTACATCGAGAGGGAATGAGCTCCCGAGC 5520
5461 AGATGGAAGAGTGTGCTTCAAACTCTTACATCGAGAGGGAATGAGCTCCCGAGC 5520
5521 AATTCAAGCAAAAGCGCTCGGTTGTGCAAAAGCGGCAACAGCAAGCGGAGGCTCTG 5580
5521 AATTCAAGCAAAAGCGGTAATCGGTTGTGCAAAAGCGGCAACAGCAAGCGGAGGCTCTG 5580
5581 CTCCGCTGTGTGAGTCCAAAGTGGCGAGCCTTGAGACCTTCTGGGCGAAGCACATGTGA 5640
5581 CTCCGCTGTGTGAAATCCAAAGTGGCGAGCCTTGAGAGCCTTCTGGGCGAAGCATATGTGA 5640
5641 ATTTCAATCAGCGGAATACGTPACTAGAGCTTTATCCACTCTCTCCCTGGGAAACCCCGGA 5700
5641 ATTTCAATCAGCGGATACAAATATTAGCAGGCTTGTCCACTCTCTCTGGCAACCCCGGA 5700
5701 TAGCATCATTTAGGCAATTTACGCTCTATCACTAGCCGCTCACCACCCAAACACCC 5760
5701 TAGCATCATCTAGGCAATTTACGCTCTATCACTAGCCGCTCACCACCCAAACATACCC 5760
5761 TCCTGTTTAAATCATCTGGGGGATGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 5820
5761 TCCTGTTTAAATCATCTGGGGGATGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 5820
5821 CAGCTTTGCTGGGCGCGGATACGCGGAGCGGCTGTGTGGAGCATAGGCTTTGGGAAG 5880
5881 TGCTGTGGAATCTTGGCGGCTATGGGAGAGGGGTAGCGGCGCATCTGTGCGCTTTA 5940
5881 TGCTGTGGAATTTTGGCGGCTATGGGAGAGGGGTAGCGGCGCGCTGTGCGCTTTA 5940
5941 AGGTCAATGAGCGGAGGTGCGCTTCCACCGGAGGAGCTGTCAACTTACTCTCTGCAATCC 6000

5941 AGGTCAATGAGCGGAGAGTGCCTCCACCGAGGAACCTGTTAACTACTCTCTGCTATCC 6000
6001 TCTCTCTGCTGCTCCCTGTTGCTGCGGCTGCTGCGGCAATACTGCTGCGGCACTGCG 6060
6001 TCTCTCTGCTGCTCCCTGTTGCTGCGGCTGCTGCGGCAATACTGCTGCGGCACTGCG 6060
6061 GCGCGGAGAGGGGCTGTGCAAGTGAACCGGCTGATAGCGTTGCTTCCGGGGTGA 6120
6061 GCGCGGAGAGGGGCTGTGCAAGTGAACCGGCTGATAGCGTTGCTTCCGGGGTGA 6120
6121 ACACGCTCTCCCTACGCACTATGCTGAGAGGAGCGCTCAGCAGCTGTCACTCAGA 6180
6121 ACACGCTCTCCCTACGCACTATGCTGAGAGGAGCGCTCAGCAGCTGTCACTCAGA 6180
6181 TCTCTCTAGCTTACCTACTCACTCACTGCTGAGGCGCTCCACAGTGTGATCAAGAG 6240
6181 TCTCTCTAGCTTACCTACTCACTCACTGCTGAGGCGCTCCACAGTGTGATCAAGAG 6240
6241 ACTGCTCTACGCACTGCTCCGGCTGTGCTGAGGAGTGTGAGATGAGATGAGATGAG 6300
6241 ACTGCTCTACGCACTGCTCCGGCTGTGCTGAGGAGTGTGAGATGAGATGAGATGAG 6300
6301 TGTGACTGACTTCAAGACCTGCTCCAGTCCAACTCTGCGGCTTACCGGAGTCC 6360
6301 TGTGACTGACTTCAAGACCTGCTCCAGTCCAACTCTGCGGCTTACCGGAGTCC 6360
6361 CTTTCTCTCATGCCAAACGCGGTACAAGGAGTGTGCGGGGGGACGCGATCATGAAA 6420
6361 CTTTCTCTCATGCCAAACGCGGTACAAGGAGTGTGCGGGGGGACGCGATCATGAAA 6420
6421 CCACCTGCGCATGCGGAGCACAGATCGCGGACATGTCAAAACGTTCCATGAGATCG 6480
6421 CCACCTGCGCATGCGGAGCACAGATCGCGGACATGTCAAAACGTTCCATGAGATCG 6480
6481 TAGGCTGTAAGCTTCACAGCAACAGCTGCGACGGAACGTTCCCATCAACGATACACCA 6540
6481 TAGGCTGTAAGCTTCACAGCAACAGCTGCGACGGAACGTTCCCATCAACGATACACCA 6540
6541 CCGGACCTTTCACACCTCCCGCGGCCAACTATTTCCAGGGGCTTATGCGGGTGGCTG 6600
6541 CCGGACCTTTCACACCTCCCGCGGCCAACTATTTCCAGGGGCTTATGCGGGTGGCTG 6600
6601 CTGAGAGTACGTGAGGTTACGCTGTGGGGATTTCCACTACGTGACGGGATACCA 6660
6601 CTGAGAGTACGTGAGGTTACGCTGTGGGGATTTCCACTACGTGAGGGGATACCA 6660
6661 CTGACAAAGTAAAGTGCCTGCTCCGCGGCCAACTATTTCCAGGGGCTTATGCGGGTGGCTG 6720
6661 CTGACAAAGTAAAGTGCCTGCTCCGCGGCCAACTATTTCCAGGGGCTTATGCGGGTGGCTG 6720
6721 GAGTGGGTTGACAGGTACGCTCCGGGCTGAAAACCTCTCTTCCAGGGGAGGATGAGT 6780
6721 GAGTGGGTTGACAGGTACGCTCCGGGCTGAAAACCTCTCTTCCAGGGGAGGATGAGT 6780
6781 TCCAGGTCGGGCTCAACCAATACCTGCTGGGTCGAGCTCCCATCGGCGCCGAAACCGG 6840
6781 TCCAGGTCGGGCTCAACCAATACCTGCTGGGTCGAGCTCCCATCGGCGCCGAAACCGG 6840
6841 ACCTAACAGTGTCTTCTCCATGCTCACGATCCCTCCCACTTACAGCAGAGAGGCTA 6900
6841 ACCTAACAGTGTCTTCTCCATGCTCACGATCCCTCCCACTTACAGCAGAGAGGCTA 6900
6901 AGCGTAGGCTGAGGCTCTCCCGCTCTTTAGCGAGCTCATGCTAGCAGCTTGT 6960
6901 AGCGTAGGCTGAGGCTCTCCCGCTCTTTAGCGAGCTCATGCTAGCAGCTTGT 6960
6961 CTGCGCTTCTTGAAGGCAACATGCACTACCGTCTATGCTCTCCCGGAGCGCTGACTCA 7020
6961 CTGCGCTTCTTGAAGGCAACATGCACTACCGTCTATGCTCTCCCGGAGCGCTGACTCA 7020
7021 TCAGGCGCAACCTCTTGTGGCGGAGAGATGGCGGGAACATCACTTCGCGTGGAGTCAG 7080
7021 TCAGGCGCAACCTCTTGTGGCGGAGAGATGGCGGGAACATCACTTCGCGTGGAGTCAG 7080

Db 9241 CCAGCTGGTTCGTTGGTTACAGCGGGAGACATATATACAGCTGTCGTGCGCC 9300
 Qy 9301 GACCCGCTGGTTCGTTGGTTACAGCGGGAGACATATATACAGCTGTCGTGCGCC 9360
 Db 9301 GACCCGCTGGTTCGTTGGTTACAGCGGGAGACATATATACAGCTGTCGTGCGCC 9360
 Qy 9361 TCCCAACCGATGAACGGGAGCTAACCACTCCAGGCGCTTAAGCC-----ATTTC 9410
 Db 9361 TCCCAACCGATGAACGGGAGCTAACCACTCCAGGCGCTTAAGCC-----ATTTC 9420
 Qy 9411 CTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9470
 Db 9421 CTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9480
 Qy 9471 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 9530
 Db 9481 TCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 9540
 Qy 9531 TGTGAAGGTCGAGCGGAGCTGACGAGAGTCTGATAGTGGCCCTCTGCGAGAT 9590
 Db 9541 TGTGAAGGTCGAGCGGAGCTGACGAGAGTCTGATAGTGGCCCTCTGCGAGAT 9600
 Qy 9591 CATGT 9595
 Db 9601 CAAGT 9605

RESULT 12
 ABK91424
 ID ABK91424 standard; DNA; 9605 BP.
 AC
 XX ABK91424;
 XX
 XX
 DT 15-NOV-2002 (first entry)
 DE Hepatitis C virus Con 1 isolate DNA mutant 1.
 XX
 KW HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
 KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
 KW internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key
 CDS Location/Qualifiers
 342..9374
 /*tag= a
 /product= "HCV polyprotein"
 /note= "The polyprotein consists of the Core, E1, E2, P7,
 NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
 mutation replace(3625,G)
 /*tag= b
 XX
 PN WO200259321-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 16-JAN-2002; 2002WO-EP000526.
 XX
 PR 23-JAN-2001; 2001US-0263479P.
 XX
 PA (RICE-) 1ST RICECHE BIOL MOLECULAR ANGELETTI.
 XX
 PI De Francesco R, Migliaccio G, Paonessa G;
 XX
 XX WPI; 2002-599793/64.
 DR
 XX New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
 PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
 PT ribosome entry site (IRES) region, useful in studying HCV replication and
 PT expression.

PS Claim 9; Page; 69pp; English.
 XX The invention relates to nucleic acid molecules comprising altered HCV
 CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
 CC internal ribosome entry site (IRES) region coding for one or more NS3,
 CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
 CC are detailed in the specification. Also included are (1) an expression
 CC vector comprising a nucleotide sequence coding for the altered nucleic
 CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
 CC recombinant cell human hepatoma cell comprising the altered nucleic acids
 CC ; (3) a recombinant cell produced by introducing into a human hepatoma
 CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
 CC replicon enhanced cell or which containing a functional HCV replicon; (5)
 CC an HCV replicon enhanced cells made in the method; and (6) measuring the
 CC ability of a compound to affect HCV activity. The HCV replicons and HCV
 CC replicon enhanced cells are useful in studying HCV replication and
 CC expression, and HCV and host cell interactions, producing HCV RNA and
 CC proteins, and providing a system for measuring the ability of a compound
 CC to modulate one or more HCV activities e.g. to discover drugs which may
 CC treat HCV mediated diseases such as liver failure, cirrhosis and
 CC hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
 CC mutant of the invention. Note: The present sequence is not shown in the
 CC specification but was created by the indexer using the HCV sequence
 CC appearing as ABK91411 and the information in Claim 9
 XX
 SQ Sequence 9605 BP; 1910 A; 2884 C; 2732 G; 2079 T; 0 U; 0 Other;
 Query Match 86.0%; Score 8248.6; DB 6; Length 9605;
 Best Local Similarity 91.3%; Pred. No. 0;
 Matches 8766; Conservative 0; Mismatches 829; Indels 10; Gaps 1;
 Qy 1 GCCAGCCCCCTGATGGGGGACACCTCCACCATGATCATCCCTGTGAGGAACTACTG 60
 Db 1 GCCAGCCCCCTGATGGGGGACACCTCCACCATGATCATCCCTGTGAGGAACTACTG 60
 Qy 61 TCTTCACGCAAGAAAGCGTCTAGCCATGGCGTGTAGTATGATGTCGTGACGCTCCAGGAC 120
 Db 61 TCTTCACGCAAGAAAGCGTCTAGCCATGGCGTGTAGTATGATGTCGTGACGCTCCAGGAC 120
 Qy 121 CCCCCCTCCCGGAGAGCCATAGTGGTCTCGGGAACCGGTGATACACCGGAATCCCGAG 180
 Db 121 CCCCCCTCCCGGAGAGCCATAGTGGTCTCGGGAACCGGTGATACACCGGAATCCCGAG 180
 Qy 181 GAGACCGGGTCTTTCTTGGATCAACCGCTCAATGCTTGGAGATTGGGGGTGCCCC 240
 Db 181 GAGACCGGGTCTTTCTTGGATCAACCGCTCAATGCTTGGAGATTGGGGGTGCCCC 240
 Qy 241 GCGAGACTGCTAGCGAGTAGTGTGGGTGCGAAAGGCGCTTGTGTACTGCTGATAGG 300
 Db 241 GCGAGACTGCTAGCGAGTAGTGTGGGTGCGAAAGGCGCTTGTGTACTGCTGATAGG 300
 Qy 301 GTGCTTSCGAGTCCCGGGAGGTCTGTAGACCGGTGACCATGAGCAGAAATCCTAAAC 360
 Db 301 GTGCTTSCGAGTCCCGGGAGGTCTGTAGACCGGTGACCATGAGCAGAAATCCTAAAC 360
 Qy 361 CTCAGAGAAAACCAACCTAACACCAACCGCGCCACAGGACGTCAGATCCCGGGCG 420
 Db 361 CTCAGAGAAAACCAACCTAACACCAACCGCGCCACAGGACGTCAGATCCCGGGCG 420
 Qy 421 GTGGTCAGATCGTTGTGGAGTTTACTGTTGCGGCGAGGGGCCAGGTTGGGTGTGC 480
 Db 421 GTGGTCAGATCGTTGTGGAGTTTACTGTTGCGGCGAGGGGCCAGGTTGGGTGTGC 480
 Qy 481 GCGGACTAGGAAGGCTTCGAGCGGTGCGAAACCTGTGTGAAGGCGACAACTATCCCAA 540
 Db 481 GCGGACTAGGAAGGCTTCGAGCGGTGCGAAACCTGTGTGAAGGCGACAACTATCCCAA 540
 Qy 541 AGGCTCCGACCCGAGGCGAGGCTTGGGCTCAGCCCGGCTACCTTGGCCCTCTATG 600
 Db 541 AGGCTCCGACCCGAGGCGAGGCTTGGGCTCAGCCCGGCTACCTTGGCCCTCTATG 600
 Qy 601 GCATGAGGCGCTGGGGTGGGCGAGGATGGCTCTGTCAACCCCGGGCTCCCGGCTAGTT 660

QY 2821 TATCTTGACCTGTGCACCACTACAAAGTGTCTTCACTAGGCTCATATGGTGGTTAC 2880
Db 2821 TACTCTTGACCTGTGCACCACTATAAGCTGTCTTCTCGTAGGCTCATATGGTGGTTAC 2880
QY 2881 AATACTTTATCACACAGACCGGAGGCGACATGCAAGTGTGGTCCCCCCCCCTCAACGTTTC 2940
Db 2881 AATATTTTATCACAGGCGCGGAGGCGACATTTGCAAGTGTGGATCCCCCCCCCTCAACGTTTC 2940
QY 2941 GGGGAGCGCGGATGCCATCATCTCTCACTGTGCGGTTCAACCCAGAGTTAATTTTGG 3000
Db 2941 GGGGGCGCGGATGCCATCATCTCTCACTGTGCGGTTCAACCCAGAGTTAATTTTGG 3000
QY 3001 ACATCACCAAACTCCTGTGCGCCATACTCGCCCGGCTCATGGTGTCTCAGGCTGSCATAA 3060
Db 3001 CCATCACCAAACTCTGTGCGCCATACTCTCGTCCACTCATGGTGTCTCAGGCTGSCATAA 3060
QY 3061 CGAGAGTCCGCTACTTCTGTGCGCGCTCAAGGCTCATCTGCTGATCATGTAGTGGCAA 3120
Db 3061 CCAAGTSCCGTACTTCTGTGCGCGCACACGGGCTCATTTCTGTCATGTCATGTGTCGGA 3120
QY 3121 AAGTCGCGGGGTCATTATGTCCTCAAAATGGTCTTTCATGAAGCTGGCGCGCTGACAGGTA 3180
Db 3121 AGTTGTCTGGGGTCATTATGTCCTCAAAATGGTCTTTCATGAAGTTGGCGCGCTGACAGGTA 3180
QY 3181 CGTACGTTTATPACCACTTTACCCCACTGCGGGACTGGGCCACCGCGGGCTTACGAGCC 3240
Db 3181 CGTACGTTTATGACCATCTCACCCCACTCGGGGACTGGGCCACCGCGGGCTTACGAGCC 3240
QY 3241 TTGCGGTGGCGTAGAGCCGCTGCTCTCTCGCCATGGAACCAAGTATACCTTGGG 3300
Db 3241 TTGCGGTGGCGTAGAGCCGCTGCTCTCTCGATATGGAGACCAAGTATATACCTTGGG 3300
QY 3301 GAGCAGACACCGCTCGCTGTGGGACATCATCTTTGGGTCTACCCGTCCTCCGCCCGGAAGGG 3360
Db 3301 GGGCAGACACCGCGGTGTGGGACATCATCTTTGGGCTCTCCGCCCGCGAGGG 3360
QY 3361 GGAAGGAGATATTTTGGGACCGGTGTATGTCCTGGAAGGGAAGGTGGCGATCTCTTG 3420
Db 3361 GGAAGGAGATATCTGGAACCGGCGAGACAGCTTTGAAGGGCAGGGTGGCGATCTCTTG 3420
QY 3421 CGCCATCACGGCTACTCCCAACAAACGCGGGGCTACTTGGTGTGATCATCTAGACC 3480
Db 3421 CGCCTATACGGCTACTCCCAACAGACGCGAGGCTACTTGGTGTGATCATCTAGACC 3480
QY 3481 TCACAGGCGGGAAGAAACAGGTTCGAAGGGAGGTCGAAGTGTCTTCTACCGCAACAC 3540
Db 3481 TCACAGGCGGGAAGAAACAGGTTCGAAGGGAGGTCGAAGTGTCTTCTACCGCAACAC 3540
QY 3541 AATCTTCTCGGACCTGCAATCAACGGCGTGTGCTGACTGTCTTACATGGCGTGGCT 3600
Db 3541 AATCTTCTCGGACCTGCGTCAATGGCGTGTGCTGACTGTCTTATCATGGTGGCGCT 3600
QY 3601 CGAAGACCTAGCGGTCCAAAAGTTCACATCACCCAAATGACACCAATGACCTAGACCTGG 3660
Db 3601 CAAGACCTTTGCGGCGCCAAAGGCCCAATCACCCAAATGACACCAATGACCTAGACCTGG 3660
QY 3661 ACCTGCTCGGTGGCAGCGCGCCCCCGGGGCGCTCCATGACACCAATGACCTAGACCTGGCA 3720
Db 3661 ACCTGCTCGGTGGCAACCGCGCCCCCGGGGCGCTTCTTGACACCAATGACCTAGACCTGGCA 3720
QY 3721 GCTCGGACCTTTAATTGCTACGAGACATGTGTATGTCATTTCCGGTGTGCGCGGAGCGG 3780
Db 3721 GCTCGGACCTTTACTTGTGTACGAGGATGCGATGTCTTCCGGTGTGCGCGGCGG 3780
QY 3781 ACAGCAGGGAGTCTACTCTCCCGGCGGCTCTCTACTGTAAGAGCTCTCTGGGTG 3840
Db 3781 ACAGCAGGGAGGACCTACTCTCCCGGCGGCTCTCTACTTGAAGGGCTCTCTGGGCG 3840
QY 3841 GTCCATGCTTTGCCCTTTCGGGCGACGCTGTGGGCGTCTTCCGGGCTGTGTGTGCAACC 3900
Db 3841 GTCCATGCTCTGCCCTTTCGGGCGACGCTGTGGGCGATCTTTCGGGCTGTGTGTGCAACC 3900
QY 3901 GGGGGTTCGGAAGCGCGTGGACTTTCATACCGCTTGAATGTAAGAACTACCATGCGGT 3960

Db 3901 GAGGGTTCGAAGCGGTGGACTTTGTACCCGTCAGTCTATGGAACACCATATGCGGT 3960
QY 3961 CTCCGCTTTCACAGACACTCAACCCCGCGCTGTACCGCAGACATTCACAGTGGCAC 4020
Db 3961 CCCCGTCTTTCAGGCAACTGTCCCTCCCGCGCTGTACCGCAGACATTCACAGTGGCCC 4020
QY 4021 ATCTGACGCTCTCTACTTGGCAGCGGCAAGAGACCAAAAGTGCCTGTGCTATGACAGCCC 4080
Db 4021 ATCTACCGCCCTACTGTGTAGCGCAAGAGGCTATAGGTGCGGCTGTGATGACAGCCC 4080
QY 4081 AAGGTAACAAGTGTCTGCTCTGAAACCGGTCTGTCGCCCACTTATAGGTTTGGGGCGT 4140
Db 4081 AAGGTAACAAGTGTCTGCTCTGAAACCGGTCTGTCGCCCACTTATAGGTTTGGGGCGT 4140
QY 4141 ATATGTCGAAGGCACAGGTCATCGACCTCAACATCAGAACTGGGTAAGGACCATTACCA 4200
Db 4141 ATATGTCGAAGGCACAGGTCATCGACCTCAACATCAGAACTGGGTAAGGACCATTACCA 4200
QY 4201 CGGCGGCTCCATTACGTACTCCACTATGGCAAGTTCCTTGCCTGCGGTGGTGTCTG 4260
Db 4201 CGGTCGCCCATCACTACTCCACTATGGCAAGTTCCTTGCCTGCGGTGGTGTCTG 4260
QY 4261 GGGCGGCTTATGACATCATTAATATGTAGTGGCACTCAACTGACTCGACTTCACTCT 4320
Db 4261 GGGCGGCTTATGACATCATTAATATGTAGTGGCACTCAACTGACTCGACTTCACTCT 4320
QY 4321 TGGGCTTCGGCACAGTCTGRCACAGCGGAGAGGCTGGAGCGGCTGCTGCTGCTG 4380
Db 4321 TGGGCTTCGGCACAGTCTGRCACAGCGGAGAGGCTGGAGCGGCTGCTGCTGCTG 4380
QY 4381 CCACCGCTACACCTCGGGATCGGTTCACGTGCGCACACCCCAATATCGAGGAATAGGCC 4440
Db 4381 CCACCGCTACGCTCGGGATCGGTTCACGTGCGCACATCCAAACATCGAGGAGTGGCTC 4440
QY 4441 TGTCCAAACATGAGAGATCCCTTCTATGGGAAAAGCATGCCATCCCATGAGGCAATAGG 4500
Db 4441 TGTCCAGACTGAGAGAAATCCCTTTTATGGCAAGCCATCCCATCGAGACATCAAGG 4500
QY 4501 GGGGGAGCATCTCATTTCTGCTCATTCAGAGAGAAATGTGACGAGCTCGCCGCAAGC 4560
Db 4501 GGGGGAGCATCTCATTTCTGCTCATTCAGAGAGAAATGTGATGAGCTCGCCGCGAGC 4560
QY 4561 TGACAGGCTTCGACTGAACTGACTGATGATATACCGGGGCTTGTATGCTCGCTCATAC 4620
Db 4561 TGTCCGCGCTCGGACTCATGCTGTAGCATATACCGGGGCTTGTATGCTCGCTCATAC 4620
QY 4621 CGCCTATCGGAGAGCTGTGTGTGGCAACAGAGCTCTAATGACGGGTTTACCGGCG 4680
Db 4621 CAATAGCGGAGAGCTCATTTGCTGTAGCAACGAGAGCTCTAATGACGGGCTTACCGGCG 4680
QY 4681 ATTTTGACTCAGTGCATGCAATACATGTGTCAACCCAGACAGTGCAGCTTTCAGCTGG 4740
Db 4681 ATTTTGACTCAGTGCATGCAATACATGTGTCAACCCAGACAGTGCAGCTTTCAGCTGG 4740
QY 4741 ATCCACCTTTCACCATTTGAGACGACCGTGCCTCCCAAGACGCGGTGTGCGCTGCAAC 4800
Db 4741 ACCCGACTTTCACCATTTGAGACGACCGTGCCTCCCAAGACGCGGTGTGCGCTGCAAC 4800
QY 4801 GGGAGGTAGAACTGGCAGGGGTAGAGTGGGATCTACAGTTTGTGACTTCAGGAGAAC 4860
Db 4801 GGGAGGTAGAACTGGTATGGGAGGATGGGATCTACAGTTTGTGACTTCAGGAGAAC 4860
QY 4861 GGCCTCGGCACTTTCGATTTCTTCTGCTGTGTGTGCTGTATGACGCGGGCTGTGCTT 4920
Db 4861 GGCCTCGGCACTTTCGATTTCTTCTGCTGTGTGTGCTGTATGACGCGGGCTGTGCTT 4920
QY 4921 GGTATGAGCTCACGCCCGCTGAGACCTCGGTTAGGTTTGGGGCTTACCTAAATACACAG 4980
Db 4921 GGTATGAGCTCACGCCCGCTGAGACCTCGGTTAGGTTTGGGGCTTACCTAAATACACAG 4980
QY 4981 GGTTCGCGCTTTCGCGAGACATCTGGAGTTCTGGGAGAGGCTTCTACAGGCTCAGCC 5040

QY 7201 TATGGGACGCGGAGCTACAAATCTTCACTGCTAGAGTCTTGGAAAGACCGGACTACG 7260
DB 7201 TATGGGACGCGGAGCTACAAATCTTCACTGCTAGAGTCTTGGAAAGACCGGACTACG 7260
QY 7261 TCCCTCCGGTGGTACACGGATGCGCATTTGCCACTTACCAAGCTTCCTCAATACACCTC 7320
DB 7261 TCCCTCCAGTGGTACACGGGTTGCTATTTGCCGCTTGCACAGGCGCTCCGATACACCTC 7320
QY 7321 CACGGAGAAAGAGACGGTTCCTGACAGAAATCCAAATGCTCTTCTGCTTGGCGGAGC 7380
DB 7321 CACGGAGAAAGAGACGGTTCCTGACAGAAATCCAAATGCTCTTCTGCTTGGCGGAGC 7380
QY 7381 TCGCCATPAGACCTTCCGGTAGCTCCGGATCGTGGCGCTTGAAGGGGACCGGACCG 7440
DB 7381 TCGCCATPAGACCTTCCGGTAGCTCCGGATCGTGGCGCTTGAAGGGGACCGGACCG 7440
QY 7441 CCCTTCCTGACTGCTGCTCCGACGAGCGGTGACAAAGGATCCGAGTTGAGTCTACTCCT 7500
DB 7441 CCCTTCCTGACTGCTGCTCCGACGAGCGGTGACAAAGGATCCGAGTTGAGTCTACTCCT 7500
QY 7501 CCATGCCCCCTTTGAAGGGAGCGGGGACCCCGATCTCAGCGACGGGTCTTTGGTCTA 7560
DB 7501 CCATGCCCCCTTTGAAGGGAGCGGGGACCCCGATCTCAGCGACGGGTCTTTGGTCTA 7560
QY 7561 CCGTAGTGAGAGCTAGTACAGATGCTGCTGCTCTCAATGCTCTATACGTGGACAG 7620
DB 7561 CCGTAAGCGAGAGGCTAGTACAGATGCTGCTGCTCTCAATGCTCTATACGTGGACAG 7620
QY 7621 GCGCCCTGATCAGCCATGCGCTGGGAGGAAAGTAACTGCTGCCATCAACCGTTGAGCA 7680
DB 7621 GCGCCCTGATCAGCCATGCGCTGGGAGGAAAGTAACTGCTGCCATCAACCGTTGAGCA 7680
QY 7681 ACTCTTTGCTCGTCAACACAACTGCTTACGCCACAACTCCGACGCGCAAGCCCTCC 7740
DB 7681 ACTCTTTGCTCGTCAACACAACTGCTTACGCCACAACTCCGACGCGCAAGCCCTCC 7740
QY 7741 GGCAGAGAAAGTCACTTTGACAGATTCGAAAGTCTCGATGATCAATACCGGAGCTAC 7800
DB 7741 GGCAGAGAAAGTCACTTTGACAGATTCGAAAGTCTCGATGATCAATACCGGAGCTAC 7800
QY 7801 TCAAGGAGATGAAGCGAGGCTCCACAGTTAAGCTAAAGCTTCTATCTATAGAGAGG 7860
DB 7801 TCAAGGAGATGAAGCGAGGCTCCACAGTTAAGCTAAAGCTTCTATCTATAGAGAGG 7860
QY 7861 CCGTGAAGCTCAGCGCCCACTTGGCCAAATCCAAATTTGGCTATGCGGCAAGGAGC 7920
DB 7861 CCGTGAAGCTCAGCGCCCACTTGGCCAAATCCAAATTTGGCTATGCGGCAAGGAGC 7920
QY 7921 TCCGGAACCTATCCAGCAGGCGCTTAACCATCCGCTCCGTTGGGAGGACTTGTGG 7980
DB 7921 TCCGGAACCTATCCAGCAGGCGCTTAACCATCCGCTCCGTTGGGAGGACTTGTGG 7980
QY 7981 AAGACACTGAAACACCAATTTGACACACCATCATGGCAAAAGTGAGTTTCTGCGTCC 8040
DB 7981 AAGACACTGAGACACCAATTTGACACACCATCATGGCAAAAGTGAGTTTCTGCGTCC 8040
QY 8041 AACCHAGAAAGGAGCGCGAAGCAGCTCGCTTATCGATTTCCAGACCTTGGAGTTC 8100
DB 8041 AACCHAGAAAGGAGCGCGAAGCAGCTCGCTTATCGATTTCCAGACCTTGGAGTTC 8100
QY 8101 GTGTATGCGAGAGATGCGCTTTACGAGCTGGTCTCCACCTTCTCAGGCGGATGG 8160
DB 8101 GTGTATGCGAGAGATGCGCTTTACGAGCTGGTCTCCACCTTCTCAGGCGGATGG 8160
QY 8161 GCTCTCTATAGGATTTCAATACTCCCCAAGCAGCGGCTGAGTTCTCTGGTGAATACCT 8220
DB 8161 GCTCTCTATAGGATTTCAATACTCCCCAAGCAGCGGCTGAGTTCTCTGGTGAATACCT 8220
QY 8221 GGAANTCAAGAAATGCGCTTATGAGCTTCTCATATGACACCGCTGTTTGACTCAACGG 8280
DB 8221 GGAANTCAAGAAATGCGCTTATGAGCTTCTCATATGACACCGCTGTTTGACTCAACGG 8280
QY 8281 TCACCTGAGAGTACATTCGTTGTAGGAGTCAATTTACCAATGTTGACTTGGCCCCCG 8340

DB 8281 TCACCTGAGATGACATCCGTTTGGAGGTCATTTACCAATGTTGACTTGGCCCCCG 8340
QY 8341 AGGCGAGACGGCCATAAGGTCGCTCACAGAGCGGCTTTACATCGGGGTCTCCCTGACTA 8400
DB 8341 AAGCCAGACAGGCGATAAGGTCGCTCACAGAGCGGCTTTACATCGGGGTCTCCCTGACTA 8400
QY 8401 ACTCAAAGGCGAGAACTGCGGTTATCGCGGTGCGCGCAAGTGGCGTCTGACGACTA 8460
DB 8401 ATTCTAAAGGCGAGAACTGCGGCTATCGCGGTGCGCGCAAGTGGCGTCTGACGACTA 8460
QY 8461 GCTCGGTAATAACCTCACAATGTTACTTGAAGGCGACTGAGCGCTTGAAGCTGCGAAGC 8520
DB 8461 GCTCGGTAATAACCTCACAATGTTACTTGAAGGCGCTGCGGCTTGAAGCTGCGAAGC 8520
QY 8521 TCCAGAGCTCCAGATGCTCGTGAACGGAGACGACCTTGTCTGTTATCTGTGAAAGCGCG 8580
DB 8521 TCCAGAGCTCCAGATGCTCGTGAACGGAGACGACCTTGTCTGTTATCTGTGAAAGCGCG 8580
QY 8581 GAACCCAGAGAGGAGCGGCGCTTACGAGCTTCAAGGAGCTATGACTAGATACTCTG 8640
DB 8581 GAACCCAGAGAGGAGCGGCGCTTACGAGCTTCAAGGAGCTATGACTAGATACTCTG 8640
QY 8641 CCCCCCGGGGATCGCCCCCAACAGATACGACTTGAAGTGAATAACATCATCTCTCT 8700
DB 8641 CCCCCCGGGGATCGCCCCCAACAGATACGACTTGAAGTGAATAACATCATCTCTCT 8700
QY 8701 CCAATGCTCAGTTCGCGACGATGCTCTGGCAAAAGGCTATCTACCTCACCCGTGACC 8760
DB 8701 CCAATGCTCAGTTCGCGACGATGCTCTGGCAAAAGGCTATCTACCTCACCCGTGACC 8760
QY 8761 CCACACCCCTTGGACGCGGCTCGGTGGGAGACAGCTAGACACATCCAGTCAATCTCT 8820
DB 8761 CCACACCCCTTGGACGCGGCTCGGTGGGAGACAGCTAGACACATCCAGTCAATCTCT 8820
QY 8821 GGCTAGGCAATATCATCATGTTATCGGCCACCTATGGGCAAGGATGTTCTGATGACTC 8880
DB 8821 GGCTAGGCAATATCATCATGTTATCGGCCACCTATGGGCAAGGATGTTCTGATGACTC 8880
QY 8881 ACTTTTCTCCATCTCTTAGCTCAAGAGCAACTTGAATAAGGCTTGGATTTGACATCT 8940
DB 8881 ATTTTCTCTCATCTCTTAGCTCAGGAACAATTTGAATAAGGCTTGGATTTGACATCT 8940
QY 8941 ACGGGGCTTCTACTTCAATGAGCCACTTACCTA CTTAGATCAATGAACGACTCCATG 9000
DB 8941 ACGGGGCTTCTACTTCAATGAGCCACTTACCTA CTTAGATCAATGAACGACTCCATG 9000
QY 9001 GTCTTAGCGCATTTACACTCCAGTTACTCTCCAGTGAATCAATAGGTTGGCTTCAT 9060
DB 9001 GCCTTAGCGCATTTTCACTCCATAGTTACTCTCCAGTGAATCAATAGGTTGGCTTCAT 9060
QY 9061 GCCTTAGGAAATTTGGGTTACCACTTGGGAACTTGGAGACATCGGGCCAGAGTGTCC 9120
DB 9061 GCCTTAGGAAATTTGGGTTACCACTTGGGAACTTGGAGACATCGGGCCAGAGTGTCC 9120
QY 9121 GCGCTAAGCTACTGTCCAGGGGGAGGCGCGCACTTGTGGCAGATACCTCTTTAACT 9180
DB 9121 GCGCTAAGCTACTGTCCAGGGGGAGGCGCGCACTTGTGGCAGATACCTCTTTAACT 9180
QY 9181 GGGCAGTAAAGCAAGCTTAAACTCACTCCAAATCCGCGCGCTCCAGCTGCACTTGT 9240
DB 9181 GGGCAGTAAAGCAAGCTTAAACTCACTCCAAATCCGCGCGCTCCAGCTGCACTTGT 9240
QY 9241 CTGCTGTTTCTGCTGTTTACAGGGGGAGACATATATCAAGCTTGTCTCTGTCGCC 9300
DB 9241 CCAGCTGTTTCTGCTGTTTACAGGGGGAGACATATATCAAGCTTGTCTCTGTCGCC 9300
QY 9301 GACCCGCTGTTTCTGCTTACTCTACTCTTTCTGTAGGGTAGGCAATTTA CTTG 9360
DB 9301 GACCCGCTGTTTCTGCTTACTCTCTACTCTTTCTGTAGGGTAGGCAATTTA CTTG 9360
QY 9361 TCCCAACCATGAACGGGAGCTTAACCACTCCAGGCTTAAGCC-----ATTTC 9410

Db 9361 TCCOCACCGATGAACGGGAGCTAAACACTCCAGGCCAATAGGCAATCCTGTTTTTTC 9420

Qy 9411 CTGTT 9470

Db 9421 CTTTTTTTTTTCTT 9480

Qy 9471 TTTTCTTTCTTTTCCCTTCTTTAATGTTGCTCCACTTAGCCCTAGTCAAGCTAGC 9530

Db 9481 TCCCTTTTTTTCCTTTCTTTTCCCTTTGTTGCTCACTTAGCCCTAGTCAAGCTAGC 9540

Qy 9531 TGTGAAGTCCGTGAGCGGCGATGCTGAGAGAGTGTGATPACGTGGCTCTCTGCAGAT 9590

Db 9541 TGTGAAGTCCGTGAGCGGCTGTGACTGAGAGAGTGTGATPACGTGGCTCTCTGCAGAT 9600

Qy 9591 CATGT 9595

Db 9601 CAAGT 9605

RESULT 13

ABK91429

ID ABK91429 standard; DNA; 9605 BP.

XX AC ABK91429;

XX DT 15-NOV-2002 (first entry)

XX DE Hepatitis C virus Con 1 isolate DNA mutant 6.

XX KW HCV; db; Con 1; adaptive mutation; liver failure; cirrhosis;

KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;

KW internal ribosome entry site; IRES; NS5A; NS5B; HCV replication; mutant.

XX OS Hepatitis C virus.

OS Synthetic.

XX FH Location/Qualifiers

FT CDS 342..9374

FT /*tag= a

FT /product= "HCV polypeptide"

FT /note= "The polypeptide consists of the Core, E1, E2, P7,

NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"

FT mutation replace(6859,C)

FT /*tag= b

XX WO200259321-A2.

XX PD 01-AUG-2002.

XX PF 16-JAN-2002; 2002WO-EP000526.

XX PR 23-JAN-2001; 2001US-0263479P.

XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

XX PI De Francesco R, Migliaccio G, Paonessa G;

XX DR WPI; 2002-599793/64.

XX New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV

PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal

PT ribosome entry site (IRES) region, useful in studying HCV replication and

PT expression.

XX Claim 9; Page: 69pp; English.

XX The invention relates to nucleic acid molecules comprising altered HCV

CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)

CC internal ribosome entry site (IRES) region coding for one or more NS3,

CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations

CC are detailed in the specification. Also included are (1) an expression

CC vector comprising a nucleotide sequence coding for the altered nucleic

CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a

CC recombinant cell human hepatoma cell comprising the altered nucleic acids

CC ; (3) a recombinant cell produced by introducing into a human hepatoma

CC cell the altered nucleic acids; (4) producing an HCV (Hepatitis C virus)

CC replicon enhanced cell or which containing a functional HCV replicon; (5)

CC an HCV replicon enhanced cells made in the method; and (6) measuring the

CC ability of a compound to affect HCV activity. The HCV replicons and HCV

CC replicon enhanced cells are useful in studying HCV replication and

CC expression, and HCV and host cell interactions, producing HCV RNA and

CC proteins, and providing a system for measuring the ability of a compound

CC to modulate one or more HCV activities e.g. to discover drugs which may

CC treat HCV mediated diseases such as liver failure, cirrhosis and

CC hepatocellular carcinoma. The present sequence is an HCV replicon Con 1

CC mutant of the invention. Note: The present sequence is not shown in the

CC specification but was created by the indexer using the HCV sequence

CC appearing as ABK91411 and the information in Claim 9

XX

SQ Sequence 9605 BP; 1910 A; 2882 C; 2733 G; 2080 T; 0 U; 0 Other;

Query Match 86.0%; Score 8248.6; DB 6; Length 9605;

Best Local Similarity 91.3%; Pred. No. 0;

Matches 8766; Conservative 0; Mismatches 829; Indels 10; Gaps 1;

Qy 1 GCCAGCCCCCTGATGGGGGCGACACTCCACCATGATCACTCCCTGTGAGGAATACTG 60

Db 1 GCCAGCCCCGATTGGGGGCGACACTCCACCATGATCACTCCCTGTGAGGAATACTG 60

Qy 61 TCTTCAGCAGAAAGCGTCTAGCCATGCGTGTAGTATGAGTGTCTGTGAGCCTCAGGAC 120

Db 61 TCTTCAGCAGAAAGCGTCTAGCCATGCGTGTAGTATGAGTGTCTGTGAGCCTCAGGAC 120

Qy 121 CCCCCCTCCCGGAGAGCCATGATGGTGTCTGGGAAACCGGTGAGTACACCGGAATTCGCAG 180

Db 121 CCCCCCTCCCGGAGAGCCATGATGGTGTCTGGGAAACCGGTGAGTACACCGGAATTCGCAG 180

Qy 181 GACGACCGGGTCTTTCTTTGATCAACCCGCTCAATGCTCGAGATTGGCGCTGCCCCC 240

Db 181 GACGACCGGGTCTTTCTTTGATCAACCCGCTCAATGCTCGAGATTGGCGCTGCCCCC 240

Qy 241 GCGAGACTGTAGCGGAGTAGTGTGGTTCGCGAAAGCCCTTGTGGTACTGCTGATAGG 300

Db 241 GCGAGACTGTAGCGGAGTAGTGTGGTTCGCGAAAGCCCTTGTGGTACTGCTGATAGG 300

Qy 301 GTGCTTGGAGTGTCCCGGAGGTCTCTGACACCGTCAACCATGAGCACGATCTTAAC 360

Db 301 GTGCTTGGAGTGTCCCGGAGGTCTCTGACACCGTCAACCATGAGCACGATCTTAAC 360

Qy 361 CTCGAGAAAAAACCAAGCTTAACACCAACCCCGCCACAGGACCTCAAGTTCGCGGCG 420

Db 361 CTCGAGAAAAAACCAAGCTTAACACCAACCCCGCCACAGGACCTCAAGTTCGCGGCG 420

Qy 421 GTGGTCAATCGTGTGGAGTTTACCTGTGCGCGAGGGGCCCGAGTGGGTGTC 480

Db 421 GTGGTCAATCGTGTGGAGTTTACCTGTGCGCGAGGGGCCCGAGTGGGTGTC 480

Qy 481 CCGGCACTAGGAAGCTTCCGAGCGGTTCGCAACCTCGTGGAAAGCGCAACATCTCCCA 540

Db 481 CCGGCACTAGGAAGCTTCCGAGCGGTTCGCAACCTCGTGGAAAGCGCAACATCTCCCA 540

Qy 541 AGGTTCGCGACCCGAGGCGAGGCTGGGTCTAGCCCGGCTACCCCTGCGGCGCTATG 600

Db 541 AGGTTCGCGACCCGAGGCGAGGCTGGGTCTAGCCCGGCTACCCCTGCGGCGCTATG 600

Qy 601 GCAATGAGGCGCTGGGTGGGAGGATGGTCTCTACCCCGCGGTCTCCGCGCTAGTT 660

Db 601 GCAATGAGGCGCTGGGTGGGAGGATGGTCTCTACCCCGCGGTCTCCGCGCTAGTT 660

Qy 661 GGGGCCCCCAGACCCCGCGGTAGGTTCGCTAACTTGGGTAAAGTTCATCATACCCCTTA 720

Db 661 GGGGCCCCCAGACCCCGCGGTAGGTTCGCTAACTTGGGTAAAGTTCATCATACCCCTTA 720

Qy 721 CATGCGGCTTCGCGATCTCATATGGGTACATTCGCTGTGCGGCGCCCCCTAGGGGCG 780

Db 721 CATGCGGCTTCGCGATCTCATATGGGTACATTCGCTGTGCGGCGCCCCCTAGGGGCG 780

2941	Db		GGGGGGCCGCGATGCGCGTCACTCCTCTCAGTCGCGAATCCACCCAGAGCTAATCTTCA	3000
3001	QY		ACATCAACAACCTCCTCGCATACTCGGCGCGCTCATGGTCTCCAGGCTGGCATAA	3060
3001	Db		CCATCACAAAATCTTGCTCGCAATCTCGTCACTCATGGTCTCCAGCTGGTAAAA	3060
3061	QY		CGAGAGTCCGCTACTTCGTGCGGCTCAAGGGCTCATTCGTGATGCATTTAGTCGAA	3120
3061	Db		CCAAAGTCCGTACTTCGTGCGGCAACAGGGCTCATTCGTGATGCATGCTGTCGGA	3120
3121	QY		AAGTCGGGGGTGCATATGTCAAATGCTTTCATGAAGCTGGGGCGCTCACAGGTA	3180
3121	Db		AGGTGCTGGGGTCATATGTCAAATGGCTCTCATGAATTTGGCGCACTCACAGGTA	3180
3181	QY		CGTACGTTTATAACCATCTTACCCTCGGGACTGGGCCAGCGGGGCTACGAGACC	3240
3181	Db		CGTACGTTTATGACCATCTACCCCACTGGGAGCTGGGCCAGCGGGGCTACGAGACC	3240
3241	QY		TTGCGGTGGCGTAGACCCGTGCTTCTCCGCAATGGAGCAAGGTCATCACCTGGG	3300
3241	Db		TTGCGGTGGCAGTTGAGCCGCTGCTTCTCATATGGAGACCAAGGTTATCACCTGGG	3300
3301	QY		GAGCAGACACGGCTGCGTGTGGGACATCATCTTGGGTCTACCGGTCTCGCCGGAAGG	3360
3301	Db		GGGCAGACACGGCGCGTGTGGGACATCATCTTGGCCCTGCCGCTCTCGCCCGCAGG	3360
3361	QY		GGAGGAGATATTTTGGGACCGGCTCATAGTCTCGAAGGGCAAGGTGGCGACTCCTTG	3420
3361	Db		GGAGGAGATACATCTGGGACCGGACAGACCTTGAAGGGAGGGTGGCGACTCCTCG	3420
3421	QY		CGCCCATACGGCCTACTCCCAACAAAGCGGGCGTACTTGGTTGCATCATCACTAGCC	3480
3421	Db		CGCCTATACGGCCTACTCCCAACAGACGCGAGGCCCTACTTGGCTGCATCATCACTAGCC	3480
3481	QY		TCACAGCCCGGACAAGAACAGAGTCGAAGGGAGGTTCAAGTGGTTTCAACGCAACAC	3540
3481	Db		TCACAGCCCGGACAGGAACAGAGTCAGGGGAGGTCCTCAAGTGGTCTCCACCGCAACAC	3540
3541	QY		AATCTTTCCTGGGACCTGCATCAACGGCTGTCTGACTGTCTACCATGGCGCTGGCT	3600
3541	Db		AATCTTTCCTGGGACCTGCCTAATGGCTGTGTGGACTGTCTATCATGGTCCGGCT	3600
3601	QY		CGAAGACCTTAGCCGCTCCAAAGGTCCAATCACCCAAATGACCAATGTAGACTCGG	3660
3601	Db		CAAAGACCTTGC CGGCCCAAGGCCCAATCACCCAAATGTACCAATGTGACGAG	3660
3661	QY		ACCTGTCGGTGGCAGGGCCCCCGGGGGCGCTCCATGACACATGCAGTGTGGCA	3720
3661	Db		ACCTGTCGGTGGCAAGCGCCCCCGGGGGCGCTTCCTTTGACACCATGCACTCGGCA	3720
3721	QY		GCTCGGACCTTACTTCGTCAGAGACATGCTGATGTCAATCCGGTGCGCCGCGAGCG	3780
3721	Db		GCTCGGACCTTACTTCGTCAGAGGATGCCGATGTCAATCCGGTGCGCCGCGGGCG	3780
3781	QY		ACAGAGGGGAAGTCTACTCTCCCGAGGCCGCTCTCTACTCAAAAGGCTCTCTCGGGTG	3840
3781	Db		ACAGAGGGGAGCTACTCTCCCGAGGCCCTCTCTACTTGAAGGGCTCTTTCGGGCG	3840
3841	QY		GTCATTTGCTTTGGCCTTGGGGCACTGTGTGGCGTCTTCGCGGCTGCTGTGTGCAACC	3900
3841	Db		GTCACATGCTCTGCGCCCTCGGGGCACTGTGGGCACTTTTCGGGCTGCGGTGTGCAACC	3900
3901	QY		GGGGGTTCGGAAGCGGTGGACTTCATCCCGTTGAGTCTATGGAACATACCATGCGGT	3960
3901	Db		GAGGGTTCGAAGCGGTGGACTTTGTACCGCTCGAGTCTATGGAACACATATGCGGT	3960
3961	QY		CTCCGCTCTTACAGACAACCTCAACCCCGCGGCTGTACCGCAGACATTCCAAGTGCAC	4020
3961	Db		CCCCGCTTTCAGACAACCTGTCCTCTCGGCGGTACCGCAGACATTCAGGTGGCCC	4020
4021	QY		ATCTGCACGCTCCTACTGGCAGCGGCAAGAGCACAAAGTGC CGGCTGCGTATGACGCC	4080

4021 ATCTACACGCCCTTACTGGTATCGGCGAAGAGCACTAAGTGTCCGGCTCGGTATCGAGCCC 4080

4081 AAGGGTACAAGGTGCTCGTCTCTGAAACCCGTCCTTCCGCCCACTTAGGGTTTGGGGCGT 4140

4081 AAGGGTAAAGGTGCTGCTCTGAAACCCGTCCTCGGCGCACCTTAGGTTTCGGGGCGT 4140

4141 ATATGTCCMAGGCAACAGGTTATCGACCTTAACATCAGATCTGGGTAAAGCAATTAACA 4200

4141 ATATGTCTTAAGGCACATCGTATCCACCTTAACATCAGAACCGGGTAAGNCCATCAACA 4200

4201 CGGGCGCTCCAACTAGCTACCTCCACTATCGGAAGTTCCTTCCGACGGTGGCTTCTG 4260

4201 CGGGTCCCCCATCAGTACTCCACTATGGCAAGTTTCTCCGACGGTGGTTCCTG 4260

4261 GGGCGCCTATGACATCATATATGTGATGAGTGCACCTCAACTGACTCGACTACCACT 4320

4261 GGGCGCCTATGACATCATATATGTGATGAGTGCACCTCAACTGACTCGACTACCACT 4320

4321 TGGGCAATCGGCAAGTCTTGACCAAGCGGAGACGGCTGGAGCGGGCTGCTGTGCTCG 4380

4321 TGCGCATCGGCACAGTCTCGACCAAGCGGAGACGGCTGGAGCGGCACTCGTCTGTGCTG 4380

4381 CCACCGCTACACTCCGGATCGGTACCGTGCACACCCCAATATCGAGGAATAGGCC 4440

4381 CCACCGCTACCGCTCCGGATCGGTACCGTGCACATCAAACATCGAGAGGTGGCTC 4440

4441 TGTCCAACTAATGGAGAGTCCCTCTATGGCAAGGCCATCCCCATTGAGGCCATCAAGG 4500

4441 TGTCCAGCACTGGAGAAATCCCTTTATGGCAAGCCATCCCATTCAGAGCAATCAAGG 4500

4501 GGGGAGGCATCTCAATTTCTGCCATTCGAAGAGAAATGTCAAGACTCCCGCAAGC 4560

4501 GGGGAGGCACTCAATTTCTGCCATTTCCAAAGAAGAAATGTGATGAGCTCGCGCGAAGC 4560

4561 TGACAGCCTCGACTGACAGCGTGTAGCATATTAACGGGGCTTGATGTCCGTCTATC 4620

4561 TGTCCGGCCTCGGACTCAATCTGTAGCATATTAACGGGGCCTTGATGTATCCGTCATC 4620

4621 CGCTATCGGAGACGTGCTGTGTGCAACAGACGCTCTAATGACGGGTTTCAACGGCG 4680

4621 CAATAGCGGAGACGTGATGTGTGTAGCAAGGACGCTCTAATAGAGGSGCTTTACGGCG 4680

4681 ATTTTGACTCAGTGTGACTGCAATACATGTGTCAACCCAGACAGTTCGACTTCAGCTTG 4740

4681 ATTTTGACTCAGTGTGACTGCAATACATGTGTCAACCCAGACAGTTCGACTTCAGCCTTG 4740

4741 ATCCGACCTTACCAATGACAGACGACCGTGCACCAAGCGGCTGTGCGCTCGCAAC 4800

4741 ACCCGACCTTCAACATGAGACGACCGTGCACCAAGACGCGGTGTACGCTCGCAGC 4800

4801 GGGAGGTAGAACTGGCAGGGGTAGGATGGCACTTACAGGTTTGTGACTCCAGAGAAC 4860

4801 GGCAGCGAGACTGGTAAGGGCAAGATGGGCATTTACAGTTTGTGACTCCAGAGAAC 4860

4861 GGCCTTCGGGCATGTTGGAATTTCTTCGTCCTGTGTGAGTGCATACGCGGGCTGTGCTT 4920

4861 GGCCTTCGGGCATGTTGGAATTTCTTCGTCCTGTGTGTGAGTGCATACGCGGGCTGTGCTT 4920

4921 GGTATGAGCTACGCCCGCTGAGACCTTCGGTTAGGTTGGGGCTTACTTAATACACAG 4980

4921 GGTATGAGCTACGCCCGCTGAGACCTTCAAGTASGTTGGGGCTTACTTAACACACAG 4980

4981 GGTTCGCTCTGCCAGAGCACTTGAGTTCTGGGAGACGCTCTTCAAGGCTCAACC 5040

4981 GGTTCGCTCTGCCAGAGCACTTGAGTTCTGGGAGACGCTTCTTACAGGCTCAACC 5040

5041 ACATAGATCCCACTTCTGTCCAGACTAAACAGCAGGAGCAACATTTGCTTACCTTGG 5100

5041 ACATAGACCCCACTTCTGTCCAGACTAAGCAGCAGGAGCAACATTTCCCTTACCTGG 5100

5101 TGGCATATCAAGCTACAGTGTGCGCCAGGCTCAAGCTCCACTCACTGTGGGACAAA 5160

5101 TAGCATACAGGCTACAGTGTGCGCCAGGCTCAGGCTCACTCACTGTGGGACAAA 5160

QY 9531 TGTGAAGGTCCTGAGCCGATGACTGCACAGAGTGTGATAGTGCCTCTCTGCAGAT 9590
 |||||
 Db 9541 TGTGAAGGTCCTGAGCCGCTTACTGTCAGAGAGTGTGATAGTGCCTCTCTGCAGAT 9600
 |||||
 QY 9591 CATCT 9595
 |||||
 Db 9601 CAAGT 9605

RESULT 14
 ABK91432
 ID ABK91432 standard; DNA; 9605 Bp.
 XX
 AC ABK91432;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Hepatitis C virus Con 1 isolate DNA mutant 9.
 XX
 KW HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
 KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
 XX internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 342..9374
 FT /*tag= a
 FT /product= "HCV polyprotein"
 FT /note= "The polyprotein consists of the Core, E1, E2, P7,
 FT NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
 FT mutation
 FT replace(6936,G)
 FT /*tag= b
 XX
 PN WO200259321-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 16-JAN-2002; 2002WO-EP000526.
 XX
 PR 23-JAN-2001; 2001US-0263479P.
 XX
 PA (RICE-) 1ST RICERCHIE BIOL MOLECOLARE ANGELETHI.
 XX
 PI De Francesco R, Migliaccio G, Paonessa G;
 XX WPI; 2002-599793/64.
 XX
 PT New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
 PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
 PT ribosome entry site (IRES) region, useful in studying HCV replication and
 PT expression.
 XX
 PS Claim 9; Page; 69pp; English.
 PS
 XX The invention relates to nucleic acid molecules comprising altered HCV
 CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
 CC internal ribosome entry site (IRES) region coding for one or more NS3,
 CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
 CC are detailed in the specification. Also included are (1) an expression
 CC vector comprising a nucleotide sequence coding for the altered nucleic
 CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
 CC recombinant cell human hepatoma cell comprising the altered nucleic acids
 CC ; (3) a recombinant cell produced by introducing into a human hepatoma
 CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
 CC replicon enhanced cell or which containing a functional HCV replicon;
 CC an HCV replicon enhanced cells made in the method; and (6) measuring the
 CC ability of a compound to affect HCV activity. The HCV replicons and HCV
 CC replicon enhanced cells are useful in studying HCV replication and
 CC expression, and HCV and host cell interactions, producing HCV RNA and
 CC proteins, and providing a system for measuring the ability of a compound

CC to modulate one or more HCV activities e.g. to discover drugs which may
 CC treat HCV mediated diseases such as liver failure, cirrhosis and
 CC hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
 CC mutant of the invention. Note: The present sequence is not shown in the
 CC specification but was created by the indexer using the HCV sequence
 CC appearing as ABK91411 and the information in Claim 9
 XX
 SQ Sequence 9605 Bp; 1911 A; 2883 C; 2732 G; 2079 T; 0 U; 0 Other;

Query Match 86.0%; Score 8248.6; DB 6; Length 9605;
 Best Local Similarity 91.3%; Pred. No. 0;
 Matches 8766; Conservative 0; Mismatches 829; Indels 10; Gaps 1;
 QY 1 GCCAGCCCCCTGATGGGGGCGACACTCCACCATGATCACTCCCTGTGAGGAATCTACTG 60
 Db 1 GCCAGCCCCCGATTGGGGGCGACACTCCACCATGATCACTCCCTGTGAGGAATCTACTG 60
 QY 61 TCTTACGCGAGAAAGCGTCTAGCCATGGCGTTAGTATGATGATGCTGTCAGAGTCTCCAGGAC 120
 Db 61 TCTTACGCGAGAAAGCGTCTAGCCATGGCGTTAGTATGATGATGCTGTCAGAGTCTCCAGGAC 120
 QY 121 CCCCCCTCCCGGAGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGGAATTCGCG 180
 Db 121 CCCCCCTCCCGGAGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGGAATTCGCG 180
 QY 181 GACGACCGGCTCTTTCTTTGGATCAACCGCTCAATGCTGAGATTTGGGGGTGCCCCC 240
 Db 181 GACGACCGGCTCTTTCTTTGGATCAACCGCTCAATGCTGAGATTTGGGGGTGCCCCC 240
 QY 241 GCGAGACTGCTAGCGAGTAGTGTGGGTGCGAAAGGCTTTGTTACTGCTCATAGG 300
 Db 241 GCGAGACTGCTAGCGAGTAGTGTGGGTGCGAAAGGCTTTGTTACTGCTCATAGG 300
 QY 301 GTGCTTGCGAGTGCCTCCGAGGCTCTCGTAGACGCTGACCATGAGCAGCAATCTTAAC 360
 Db 301 GTGCTTGCGAGTGCCTCCGAGGCTCTCGTAGACGCTGACCATGAGCAGCAATCTTAAC 360
 QY 361 CTCAAGAAAAAACCAACCTAACCAACCGCGGCCACAGGACCTCAAGTTCCCGGGCG 420
 Db 361 CTCAAGAAAAAACCAACCTAACCAACCGCGGCCACAGGACCTCAAGTTCCCGGGCG 420
 QY 421 GTGCTCAGATCTTGTGGAGTTTACCTGTTGCCGCGAGGGGCCCGAGGTGGGTGTGC 480
 Db 421 GTGCTCAGATCTTGTGGAGTTTACCTGTTGCCGCGAGGGGCCCGAGGTGGGTGTGC 480
 QY 481 GCGGACTAGGAAGGCTTCCGAGGCGTCCGACCTCTGCGAGGCGACCACTATCCCA 540
 Db 481 GCGGACTAGGAAGGCTTCCGAGGCGTCCGACCTCTGCGAGGCGACCACTATCCCA 540
 QY 541 AGGCTGCGCGACCCCGAGGCGAGGGCTTGGGCTCAGCCCGGTTACCTTGGCCCTCTATG 600
 Db 541 AGGCTGCGCGACCCCGAGGCGAGGGCTTGGGCTCAGCCCGGTTACCTTGGCCCTCTATG 600
 QY 601 GCATGAGGGCTTGGGGTGGGAGATGCTCTGTTACCCCGGGCTCCCGGGCTCCCGGCTAGTT 660
 Db 601 GCATGAGGGCTTGGGGTGGGAGATGCTCTGTTACCCCGGGCTCCCGGGCTCCCGGCTAGTT 660
 QY 661 GGGGCCCCACGACCCCGCGGTAGTCTCGCTAACTTGGGTAAGTCAATCATATACCTTTA 720
 Db 661 GGGGCCCCACGACCCCGCGGTAGTCTCGCTAACTTGGGTAAGTCAATCATATACCTTTA 720
 QY 721 CATCGGCTTCCCGATCTCATGGGGTACATTCGCTGTCGCGGCGCCCGCTAGGGGCG 780
 Db 721 CATCGGCTTCCCGATCTCATGGGGTACATTCGCTGTCGCGGCGCCCGCTAGGGGCG 780
 QY 781 CTGCGAGGGCTTGGCAGACCGGTCTCGGGTCTGCGAGGACGGGTGAATATGCAACAG 840
 Db 781 CTGCGAGGGCTTGGCAGACCGGTCTCGGGTCTGCGAGGACGGGTGAATATGCAACAG 840
 QY 841 GGAACTTCCCGGTTGCTTTCTTCTATCTTCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCT 900
 Db 841 GGAACTTCCCGGTTGCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 900

QY 901 TCCAGCTCCGCTTATGAAGTGGCAACGCTGTCGGGATATACCATGTCAGAACGACT 960
 Db 901 TCCAGCTCCGCTTATGAAGTGGCAACGCTGTCGGGATATACCATGTCAGAACGACT 960
 QY 961 GTTCCAACTCAAGCATGTGTATGAGGACGCGATGATCATATCTCCCGGCTGG 1020
 Db 961 GTTCCAACTCAAGCATGTGTATGAGGACGCGATGATCATATCTCCCGGCTGG 1020
 QY 1021 TCCCTGTGTTTCCAGAGGTAACAGCTCCGCTGTGTGGTAGCGTCACTCCACGCTCG 1080
 Db 1021 TCCCTGTGTTTCCAGAGGTAACAGCTCCGCTGTGTGGTAGCGTCACTCCACGCTCG 1080
 QY 1081 CGGCCAGGAATGCCAGCTGCCCATACGACAAATACGACGCCAGCTGCACTGCTGTTG 1140
 Db 1081 CGGCCAGGAATGCCAGCTGCCCATACGACAAATACGACGCCAGCTGCACTGCTGTTG 1140
 QY 1141 GGACGGCTGCTTCTGCTCCGCTATGATAGTGGGGATCTCTGCGGATCTATTTTCCCTCG 1200
 Db 1141 GGACGGCTGCTTCTGCTCCGCTATGATAGTGGGGATCTCTGCGGATCTATTTTCCCTCG 1200
 QY 1201 TCTCCAGCTGTTCACCTTCTGCTCGCGGATGAGACAGTGCAGGACTGCAACTGCT 1260
 Db 1201 TCGCCAGCTGTTCACCTTCTGCTCGCGGATGAGACAGTGCAGGACTGCAACTGCT 1260
 QY 1261 CAATCTATCCCGGCTATGATCAGGTCAACGATGCTGCTGGGATATGATGATGATGATGAT 1320
 Db 1261 CAATCTATCCCGGCTATGATCAGGTCAACGATGCTGCTGGGATATGATGATGATGATGAT 1320
 QY 1321 CACCTAACAGCCCTAGTGTGTCGAGTGTCTCGGATGCTCCGATCCCAAGCTGCTGGACA 1380
 Db 1321 CACCTAACAGCCCTAGTGTGTCGAGTGTCTCGGATGCTCCGATCCCAAGCTGCTGGACA 1380
 QY 1381 TGGTGGCGGGGCCCACTGGGGAGTCTGGGGGCTTGCCTACTATTCATGCTAGGGA 1440
 Db 1381 TGGTGGCGGGGCCCACTGGGGAGTCTGGGGGCTTGCCTACTATTCATGCTAGGGA 1440
 QY 1441 ACTGSCATAAGTGTGATGTTGGGCTACTCTTTCCGGGCTGTAACGGGAGACCCACA 1500
 Db 1441 ACTGSCATAAGTGTGATGTTGGGCTACTCTTTCCGGGCTGTAACGGGAGACCCACA 1500
 QY 1501 CGACGGGAGGCTGCGCGGCCACACCTCCGCGTTCACGCTCCCTTTCTCATCTGGG 1560
 Db 1501 TGAAGGGGGACGATGCGCAAAACCTCCGGATTTACGCTCCCTTTTACCCGGGT 1560
 QY 1561 CGTCTCAGAAATCAAGCTGTGATATACCAACGAGCTGCAATCAACAGGACTGCC 1620
 Db 1561 CATCCAGAAATCAAGCTGTGATATACCAACGAGCTGCAATCAACAGGACTGCC 1620
 QY 1621 TAAATGCAATGACTCCCTCCAACTGGGTTCTTTGCGGCTGTTTACGCAACAGT 1680
 Db 1621 TGAATGCAATGACTCCCTCCAACTGGGTTCTTTGCGGCTGTTTACGCAACAGT 1680
 QY 1681 TCAACTGCTCCGGTGGCGGAGCGCATGCGAGCTGCGCGCCCATTTGACTGGTTCCGCC 1740
 Db 1681 TCAACTCATCTGGATGCCAGAGCGCATGCGAGCTGCGAGCCCATCGACCGTTCCGTC 1740
 QY 1741 AGGGTGGGGCCCATCACTCATATTAAGCTAAGCTAAGCTCGGATCGAGAGCTTTATTGCT 1800
 Db 1741 AGGGTGGGGCCCATCACTCATATTAAGCTAAGCTAAGCTCGGATCGAGAGCTTTATTGCT 1800
 QY 1801 GGCAATTACGCGCTCGACCGTGTGCTGCTGACCGGCTGCGAGGTGTGCTCCAGT 1860
 Db 1801 GGCAATTACGCGCTCGACCGTGTGCTGCTGACCGGCTGCGAGGTGTGCTCCAGT 1860
 QY 1861 ATTGTTTCAACCCCAAGCCCTGTTGTGTTGGGACACACGATGCTTCCGGTGTCCCTACGT 1920
 Db 1861 ACTGCTTCAACCCCAAGCCCTGTTGTGTTGGGACACACGATGCTTCCGGTGTCCCTACGT 1920
 QY 1921 ATAGCTGGGGGAGAAATGAGACAGACGTGATGCTTCTCAACCAACACGCTGCCGACAA 1980
 Db 1921 ACAGTGGGGGAGAAATGAGACAGACGTGATGCTTCTTAAACCAACACGCTGCCGACAA 1980
 QY 1981 GCAACTGGTTCGGCTGTACATGGATGAATAGTACTGGGTTCACTAAGACGCTGGAGGTC 2040

1981 GCAACTGGTTCGGCTGTACATGGATAGTACCTGGGTTCAACAGACGTCGCGGGGCC 2040
 2041 CCCGCTGTAAACATCGGGGGGTGCGTAACCGACCTTGANTGCGCCCAACGAGCTCTCC 2100
 2041 CCCGCTGTAAACATCGGGGGGTGCGTAACCGACCTTGANTGCGCCCAACGAGCTCTCC 2100
 2101 GGAGACACCCGAGGCTACTTACACAAAATGTGGCTGGGGCCCTGTTGACACTAGT 2160
 2101 GGAGACACCCGAGGCTACTTACACAAAATGTGGCTGGGGCCCTGTTGTTGACACCAAT 2160
 2161 GCCTAGTAGACTACCCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTTCATCT 2220
 2161 GCCTAGTAGACTACCCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTTCATCT 2220
 2221 TTAGGTTAGATGATGTGGGGGCTGAGACAGAGCTCAATGCGCATGMAATTTGGA 2280
 2221 TCAAGGTTAGATGATGTGGGGGCTGAGACAGAGCTCAATGCGCATGMAATTTGGA 2280
 2281 CTCGAGGAGAGCGCTGTAACTTGGAGACAGGATAGTCAAGTCAAGCCCGCTGCTGC 2340
 2281 CTCGAGGAGAGCGCTGTGTAACTTGGAGACAGGATAGTCAAGTCAAGCCCGCTGCTGC 2340
 2341 TGTCTAACAGAGTGGCAGATCTGCTGCTTTCACACCTTACCCGCTTTATCCA 2400
 2341 TGTCTAACAGAGTGGCAGATCTGCTGCTTTCACACCTTACCCGCTTTATCCA 2400
 2401 CTGCTTGTATCCATCTCCATCAGAACATCTGGAGTACATCTGTTGCTTTTCTTCTCTCG 2460
 2401 CTGCTTGTATCCATCTCCATCAGAACATCTGGAGTACATCTGTTGCTTTTCTTCTCTCG 2460
 2461 CAGCGTTGCTCTCTTTCGCAATCAATGGGATACATCTGTTGCTTTTCTTCTCTCTCG 2520
 2461 CAGCGTTGCTCTCTTTCGCAATCAATGGGATACATCTGTTGCTTTTCTTCTCTCTCG 2520
 2521 CAGACGCGCGCTGTGCTGCTTGTGGATGATGCTGTGCTGTGATAGCTCAAGCTGAGCGG 2580
 2521 CAGACGCGCGCTGTGCTGCTTGTGGATGATGCTGTGCTGTGATAGCTCAAGCTGAGCGG 2580
 2581 CTTTGTAGAACTTGTGCTGCTTGTGGATGATGCTGTGCTGTGATAGCTCAAGCTGAGCGG 2640
 2581 CTTTGTAGAACTTGTGCTGCTTGTGGATGATGCTGTGCTGTGATAGCTCAAGCTGAGCGG 2640
 2641 CTTTGTGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
 2641 CTTTGTGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
 2701 CGTATGCTTTTATGCGCTATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760
 2701 CATATGCTTCTACGCGCTATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760
 2761 CTTTACGCTTGGACCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820
 2761 CATACGCTATGAGACCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820
 2821 TATTTTGTGCTTGTCAACATACTACAAAGTGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880
 2821 TATTTTGTGCTTGTCAACATACTACAAAGTGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880
 2881 AATCTTTATACAGAGCGGAGCGCATATGAGTGTGGGTCCTGCTGCTGCTGCTGCTGCTGCTGCT 2940
 2881 AATCTTTATACAGAGCGGAGCGCATATGAGTGTGGGTCCTGCTGCTGCTGCTGCTGCTGCTGCT 2940
 2941 GGGGAGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000
 2941 GGGGAGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000
 3001 ACATCAACAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060
 3001 ACATCAACAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060
 3061 CGAGAGTGGCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3120

Db	3061	CGAAGTCGCGTACTTTGTCGCGCGACACGGECTCATTTCTGTCATGCAATGCTGTGTCGGGA	3120
Qy	3121	AAGTCGCGCGGGTCATTATGTCACAAATGGTCTTTCATGAAGCTCGGCGCGCTGCACAGTA	3180
Db	3121	AGGTTGCTGGGGTCATTATGTCACAAATGGTCTTTCATGAAGTTGGCCGCACTGCACAGTA	3180
Qy	3191	CGTAGCTTTATAACCACTTTACCCCACTTGGCGGACTGGGCCACGCGGGGCTTACGAGACC	3240
Db	3181	CGTAGCTTTATAACCACTTACCCCACTTGGCGGACTGGGCCACGCGGGGCTTACGAGACC	3240
Qy	3241	TTGCGGTGCGGTAGAGCCGTCCTCTTCTCCGCCATGAGACCAAGGTATCATCCTGGG	3300
Db	3241	TTGCGGTGCGGTAGAGCCGTCCTCTCTCTGAATGAGACCAAGGTATCATCCTGGG	3300
Qy	3301	GAGCAGACACCGCTCGCTGTGGGACATCATTTGGGCTTACCGTCTCCGCCCGGAAGGG	3360
Db	3301	GGGAGACACCGCGCGGTGTGGGACATCATTTGGGCTTCCCGTCTCCGCCCGGAAGGG	3360
Qy	3361	GGAGGAGATATTTTGGGACCGGCTGATAGTCTCGAAGGCAAGGTGCGCATCTCTTG	3420
Db	3361	GGAGGAGATACATCTGGGACCGGACACAGCTTGAAGGCGAGGGGTGGCGACTCTCTG	3420
Qy	3421	CGCCCATCACGGCTACTTCCCAACAAACGGGGCGGTACTTGGTGCATCATCCTAGCC	3480
Db	3421	CGCCCTATACGGCTACTTCCCAACAGACGGAGGCTACTTGGTGCATCATCCTAGCC	3480
Qy	3481	TCACAGCCGGGACAGAAACAGGTCTGAAGGGAGGTTCAAGTGGTTTCTACCGCAACAC	3540
Db	3481	TCACAGCCGGGACAGGAACAGGTCTGAAGGGAGGTTCAAGTGGTCTCCACCGCAACAC	3540
Qy	3541	AATCTTTCTGGCACTGTGATCAACCGCGTGTCTGCACTGTCTACCATGGCGCTGGCT	3600
Db	3541	AATCTTTCTGGCACTGTGATCAATGGCGTGTGTGCACTGTCTATCATGGTCCGGCT	3600
Qy	3601	CGAAGACCTTAGCGGTCCAAAGGTCCATCCACCAAAATGTACACATGTAGACTGTG	3660
Db	3601	CAAAAGACCTTTGCGGGCCAAAGGGCCCAATCAACCAATGTACACAAATGTGACGAG	3660
Qy	3661	ACCTCGTGGCTGCGAGCGCCCCCGGGCGGCTCCATGACACCATGCACTGTGGCA	3720
Db	3661	ACCTCGTGGCTGCGAGCGCCCCCGGGCGGCTGCTTGACACCATGCACTGTGGCA	3720
Qy	3721	GCTCGGACCTTTACTTGGTCAAGACAATGCTGATGTCAATCCGGTGCGCCGCGAGGG	3780
Db	3721	GCTCGGACCTTTACTTGGTCAAGGACATGCCGATGTCAATCCGGTGCGCCGCGGGCG	3780
Qy	3781	ACAGAGGGGAAGTCTACTCTCCCGAGCGCGCTCTCTACTCTGAAAGGTCTCTCGGTG	3840
Db	3781	ACAGCAGGGGAGCCTACTCTCCCGAGCGCGCTCTCTACTCTTGAAGGGGTCTCTCGGGC	3840
Qy	3841	GTCCATGCTTTGCGCTTCGGGCGACATGCTGGGCGTCTTCCGGGTGCTGTGTGACCC	3900
Db	3841	GTCCACTGTCTGCGCTTCGGGCGACGCTGTGGGCATCTTTCCGGGTGCGGTGTGACCC	3900
Qy	3901	GGGGGTGCGAAGCGGTGGACTTCATACCGGTGAGTCTATGGAATATGCAATCCGT	3960
Db	3901	GAGGGGTGCGAAGCGGTGGACTTTGTACCGGTGAGTCTATGGAATATGCAATCCGT	3960
Qy	3961	CTCCGCTTTCACAGACAACTCAACACCCCGGCTGTACGGCAGACATTTCCAAGTGGAC	4020
Db	3961	CCCCTGCTTTCAGGACAACTCGTCCCTTCGGCGGTACCGCAGACATTTCAAGTGGCC	4020
Qy	4021	ATCTGACGCTCTACTGCGAGCGGCAAGACACCAAGTCCGGTGGGTATGACAGCC	4080
Db	4021	ATCTACAGCCCTACTGTGTAGCGGCAAGACACTAAGTTCGGGTGCGGTATGACAGCC	4080
Qy	4081	AAGGTTACAGGTGCTCGTCTGAAACCGGTGCTGTGCGGCACTTAGGTGTTGGGGCT	4140
Db	4081	AAGGTTATAAGGTGCTGTCTGAAACCGGTGCTGTGCGGCACTTAGGTGTTGGGGCT	4140
Qy	4141	ATATGTTCCAAAGGCACAGGTATCGACCTTAACATCAGAACTGGGTGAAGCACTATCA	4200
Db	4141	ATATGTTCAAGGCACATGGTATCGACCTTAACATCAGAACTGGGTGAAGCACTATCA	4200

Qy	4201	CGGGCGGCTCCATTAACGTACTCCACCTATGGCAAGTTCTTTCGCGACGGTGGCTGTCTCTG	4260
Db	4201	CGGGTGCCTCCCATCAGCTACTCCACCTATGGCAAGTTCTTTCGCGACGGTGGCTGTCTCTG	4260
Qy	4261	GGGGCGCTATGACATCATATATAATGATAGATGCCACTCAATGACTCGACTACCATCT	4320
Db	4261	GGGGCGCTATGACATCATATATAATGATAGATGCCACTCAATGACTCGACTACCATCTATCC	4320
Qy	4321	TGGGCAATGGCACAGTCTTGACCAAGCGAGACGGCTGGAGGCGGCTCGTCTGCTCTCG	4380
Db	4321	TGGGCATCGGCACAGTCTTGACCAAGCGAGACGGCTGGAGGCGGCTCGTCTGCTCTCG	4380
Qy	4381	CCACGGCTACACCTCCGGATCGGTTACCGTGCCACACCCCAATATCGAGGAATATAGGCC	4440
Db	4381	CCACGGCTACGCTCCGGATCGGTCACCGTGCCACATCCAAACATCGAGGAGGTGGCTC	4440
Qy	4441	TGTCCAAACATGGAGAGATCCCTTCTATATGGCAAGCCATCCCCATATGAGGCCATCAAGG	4500
Db	4441	TGTCCAGCACTGGAGAAATCCCTTTTATGGCAAGCCATCCCCATCGAGACCATCAAGG	4500
Qy	4501	GGGGGAGGCATCTCATTTTCTGCCATTCGAAGAAGAAATGTGACGAGCTCGCCGGAAGC	4560
Db	4501	GGGGGAGGCACCTCATTTTCTGCCATTCGAAGAAGAAATGTGATGAGCTCGCCGCGAAGC	4560
Qy	4561	TGACAGGCTCGGACTCAAACGCTGTAGCATATATACGGGSCCTTGATGTTCGTCATAC	4620
Db	4561	TGTCGGCTCTGGACTCAATGCTGTAGCATATATACGGGSCCTTGATGTATCCGTCATAC	4620
Qy	4621	CGCTATCGGAGACGTGGTTGTGTGTGGCAACAGACGCTCTAATGACGGGTTTACCGCGC	4680
Db	4621	CAACTAGCGGAGACGTCAATTTGTGTGACACGACGCTCTAATGACGGGCTTTACCGGCG	4680
Qy	4681	ATTTTGACTCAGTGATCGACTGCAATACATGTCACCCAGACAGTCGACTTCAGCTTGG	4740
Db	4681	ATTTGACTCAGTGATCGACTGCAATACATGTCACCCAGACAGTCGACTTCAGCTTGG	4740
Qy	4741	ATCCCACTTTCAACATTTGACAGCACACCGTGCCCAAGACGGGTTTCGCGTTCGCAAC	4800
Db	4741	ACCCGACCTTCAACATTTGACAGCACACCGTGCCCAAGACGGGTTTCACGCTCGCAGC	4800
Qy	4801	GGCAGGTAGAACTGGCAGGGTAGAGTGGCATCTACAGTTTGTGACTCCAGGAGAAC	4860
Db	4801	GGCAGGACGACATGGTAGGGCAGATGGGCATTTACAGTTTGTGACTCCAGGAGAAC	4860
Qy	4861	GGCCCTCGGGCATGTTTCGATCTTTCGGTCTGCTGTGAGTGTCTATGACGGGGCTGCTT	4920
Db	4861	GGCCCTCGGGCATGTTTCGATCTTTCGGTCTGCTGTGAGTGTCTATGACGGGCTGTGCTT	4920
Qy	4921	GGTATGAGCTTCAAGCCGCTTGACACCTCGGTTAGGTTGCGGGCTTACCTAATACACAG	4980
Db	4921	GGTATGAGCTTCAAGCCGCTTGACACCTCGGTTAGGTTGCGGGCTTACCTAATACACAG	4980
Qy	4981	GGTTGCCCGCTCTGCAGGACCACTGGAGTTCTGGAGAGCGTCTTTCACAGGCGCTCACCC	5040
Db	4981	GGTTGCCCGCTCTGCAGGACCACTGGAGTTCTGGAGAGCGTCTTTCACAGGCGCTCACCC	5040
Qy	5041	ACATAGTGGCCATCTTCGTCACCACTAAACAGGACGACAACTTTCCCTTACCTGG	5100
Db	5041	ACATAGACGGCCATTTCTTCTCCAGACTTAAGCAGGACGAGACAACTTCCCTTACCTGG	5100
Qy	5101	TGGCATATCAAGCTACAGTGTGGCCAGAGGCTCAAGCTCCACTCATCTGTGGGACCAAA	5160
Db	5101	TAGCATACAGGCTACGTTGTGGCCAGGGCTCAGGCTCCACTCCATCTGTGGGACCAAA	5160
Qy	5161	TGTGGAAGTGTCTCATAGGCTTAAAGCCTTAAGCTGACGGGCGCAACCCCTGTGTATA	5220
Db	5161	TGTGGAAGTGTCTCATAGGCTTAAAGCCTTAAGCTGACGGGCGCAACCCCTGTGTATA	5220
Qy	5221	GGCTAGGAGCGCTCCAAATAGGTGATCTCTCACACACCCCACTACTAAATATCATATGG	5280
Db	5221	GGCTAGGAGCGCTTCAAAACAGGTTTACTACACACACCCCACTAACCAAAATATCATATGG	5280

QY 5281 CATGCAATGTCGGTGAACCTGGAGTCTCACTAGCACTTGGGTCTGTAGCGGAGTCC 5340
 Db 5281 CATGCAATGTCGGTGAACCTGGAGTCTCACTAGCACTTGGGTCTGTAGCGGAGTCC 5340
 QY 5341 TTGCACTTTGGCGCATATGCTGACAGCAGGAGTGTGTCTATTGGGAGGATCA 5400
 Db 5341 TTGCACTTTGGCGCATATGCTGACAGCAGGAGTGTGTCTATTGGGAGGATCA 5400
 QY 5401 TCTTTCGGGAGCGAGTGTGGTTCGGGACAGGGAAGTCTCTACCAAGGAGTTCGATG 5460
 Db 5401 TCTTTCGGGAGCGAGTGTGGTTCGGGACAGGGAAGTCTCTACCAAGGAGTTCGATG 5460
 QY 5461 AGATGGAAGAGTGTGCTCACTTCCCGACAGGGAAGTCTCTACCAAGGAGTTCGATG 5520
 Db 5461 AGATGGAAGAGTGTGCTCACTTCCCGACAGGGAAGTCTCTACCAAGGAGTTCGATG 5520
 QY 5521 AATTCAAGCAAAAGGCGTTCGGGTGTGTCGAACAGGCGCATCTTCTGGGCGAAGCACAATGTGA 5580
 Db 5521 AATTCAAGCAAAAGGCGTTCGGGTGTGTCGAACAGGCGCATCTTCTGGGCGAAGCACAATGTGA 5580
 QY 5581 CTCCCGTGTGGAGTCCAAAGTGGCGAGCCCTTTCGAGACCTTCTGGGCGAAGCACAATGTGA 5640
 Db 5581 CTCCCGTGTGGAGTCCAAAGTGGCGAGCCCTTTCGAGACCTTCTGGGCGAAGCACAATGTGA 5640
 QY 5641 ATTTCATCAGCGGAATACAGTACCTAGCAGGCTTATCCACTCTGCTGTAACACCCCGGA 5700
 Db 5641 ATTTCATCAGCGGAATACAGTACCTAGCAGGCTTATCCACTCTGCTGTAACACCCCGGA 5700
 QY 5701 TAGCATATTCATGCAATTTACAGCTTCTATCACTAGCCCTTACACCCAGCAGGCTGCGT 5760
 Db 5701 TAGCATATTCATGCAATTTACAGCTTCTATCACTAGCCCTTACACCCAGCAGGCTGCGT 5760
 QY 5761 TCCCTGTTTAACTTTGGGGGATGGTGGTGGCCCACTGCTCTCCAGCGCTGGT 5820
 Db 5761 TCCCTGTTTAACTTTGGGGGATGGTGGTGGCCCACTGCTCTCCAGCGCTGGT 5820
 QY 5821 CAGCTTTGCTGGCGCGCATCGCGAGCGCTGTGGAGGATAGGCTGGGAAG 5880
 Db 5821 CAGCTTTGCTGGCGCGCATCGCGAGCGCTGTGGAGGATAGGCTGGGAAG 5880
 QY 5881 TGCTGCTGGAATTTGGCGGGCTATGGGCGAGGCTAGCGGCGCATCTGCTGGCCTTTA 5940
 Db 5881 TGCTGCTGGAATTTGGCGGGCTATGGGCGAGGCTAGCGGCGCATCTGCTGGCCTTTA 5940
 QY 5941 AGTCAATGAGCGGAGTGGCTCCACCGAGGACCTGCTCACTTACTCTGCTGCTATCC 6000
 Db 5941 AGTCAATGAGCGGAGTGGCTCCACCGAGGACCTGCTCACTTACTCTGCTGCTATCC 6000
 QY 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060
 Db 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060
 QY 6061 GCCCGGAGAGGGGCTGTGAGTGGATGAACCGGCTGATAGCTTCTGCTGCGGGGTA 6120
 Db 6061 GCCCGGAGAGGGGCTGTGAGTGGATGAACCGGCTGATAGCTTCTGCTGCGGGGTA 6120
 QY 6121 ACCAGCTCTCCCTTAGCGATGATGCTGAGCGAGCGTGTGAGCGAGTGTCTACTCAGA 6180
 Db 6121 ACCAGCTCTCCCTTAGCGATGATGCTGAGCGAGCGTGTGAGCGAGTGTCTACTCAGA 6180
 QY 6181 TCTCTCTAGCTTACCATCACTCAACTGCTGAGCGGCTCCACAGTGGATTAATGAGG 6240
 Db 6181 TCTCTCTAGCTTACCATCACTCAACTGCTGAGCGGCTCCACAGTGGATTAATGAGG 6240
 QY 6241 ACTGCTCTAGCTTACCATCACTCAACTGCTGAGCGGCTCCACAGTGGATTAATGAGG 6300
 Db 6241 ACTGCTCTAGCTTACCATCACTCAACTGCTGAGCGGCTCCACAGTGGATTAATGAGG 6300
 QY 6301 TGTTCATGCTTCAAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6360
 Db 6301 TGTTCATGCTTCAAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6360
 QY 6361 CTTTCCCTGATGCAACCGGGGTACAGGGAGTCTGGCGGGGAGCGGATCATATGCAAA 6420

Db 6361 CTTTCTTCTCACTGTCACCTGGGTACAGGGAGTCTGGCGGGGACGCGATCATATGCAAA 6420
 QY 6421 CCACCTGCCCCATGCGGAGCAGAGTTCGGGACATGTCAAAAACGTTTCCATGAGATCG 6480
 Db 6421 CCACCTGCCCCATGCGGAGCAGAGTTCGGGACATGTCAAAAACGTTTCCATGAGATCG 6480
 QY 6481 TAGGSCCTTAGAACCTGACAGCAGCAGTGTGCAAGCCTTCCCAATCAACGATACCA 6540
 Db 6481 TAGGSCCTTAGAACCTGATTAACACGTTGCGATGGAACATTTCCCATTAACGATACCA 6540
 QY 6541 CGGACCTTTCACACCTTCCCGGCGCCACTATTCAGGCGCTATGCGGGTGGCTG 6600
 Db 6541 CGGACCTTTCACACCTTCCCGGCGCCACTATTCAGGCGCTATGCGGGTGGCTG 6600
 QY 6601 CTGAGGAGTACCTGAGGTTACCGGCTGGGGATTTCCACTACGTTGACGGGATGACCA 6660
 Db 6601 CTGAGGAGTACCTGAGGTTACCGGCTGGGGATTTCCACTACGTTGACGGGATGACCA 6660
 QY 6661 CTGACAAAGTAAAGTGGCCATGCGAGGTTCCGGGCGCCGAAATTTCTTACGAGGTGATG 6720
 Db 6661 CTGACAAAGTAAAGTGGCCATGCGAGGTTCCGGGCGCCGAAATTTCTTACAGAGTGGATG 6720
 QY 6721 GAGTCCGTTTGCACAGGTACGCTCGGCGTGCACAACTTCTTACGGAGGAGCTCACT 6780
 Db 6721 GAGTCCGTTTGCACAGGTACGCTCGGCGTGCACAACTTCTTACGGAGGAGCTCACT 6780
 QY 6781 TCCAGTCCGGCTCAACAAATCTTGTGCGGTGCGAGCTCCCAATGCGAGCGCAACCGG 6840
 Db 6781 TCCAGTCCGGCTCAACAAATCTTGTGCGGTGCGAGCTCCCAATGCGAGCGCAACCGG 6840
 QY 6841 ACCTAAACAGTCTTACCTTCCAGGCTACCGATCCCTCCCACTTACAGCAGAGCAGCTA 6900
 Db 6841 ACCTAAACAGTCTTACCTTCCAGGCTACCGATCCCTCCCACTTACAGCAGAGCAGCTA 6900
 QY 6901 AGCTAGGCTGGCTAGAGGCTCTCCCTCTTCTTACGAGCTCATAGTAGCCAGTTGT 6960
 Db 6901 AGCTAGGCTGGCTAGAGGCTCTCCCTCTTCTTACGAGCTCATAGTAGCCAGTTGT 6960
 QY 6961 CTGCGCTTCTTGAAGGCAACATGACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7020
 Db 6961 CTGCGCTTCTTGAAGGCAACATGACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7020
 QY 7021 TCGAGGCAACCTTCTGCGGCGAGATGGCGGGAACATCACCTGCTGCTGCTGCTGCTGCTGCT 7080
 Db 7021 TCGAGGCAACCTTCTGCGGCGAGATGGCGGGAACATCACCTGCTGCTGCTGCTGCTGCTGCT 7080
 QY 7081 AGAATAAGTGTAGTAATTTCTGGAATCTTTCGAAACCGCTTTCAGCGGAGGAGTACAGAGG 7140
 Db 7081 AGAATAAGTGTAGTAATTTCTGGAATCTTTCGAAACCGCTTTCAGCGGAGGAGTACAGAGG 7140
 QY 7141 AGATATCCGTCGCGGCGAGATCTCGGAAATTCAGAGAGTTCCTCTGAGGAGTTCCTCTGAGG 7200
 Db 7141 AGATATCCGTCGCGGCGAGATCTCGGAAATTCAGAGAGTTCCTCTGAGGAGTTCCTCTGAGG 7200
 QY 7201 TATGGGCAACCTTCTGCGGCGAGATCTCGGAAATTCAGAGAGTTCCTCTGAGGAGTTCCTCTGAGG 7260
 Db 7201 TATGGGCAACCTTCTGCGGCGAGATCTCGGAAATTCAGAGAGTTCCTCTGAGGAGTTCCTCTGAGG 7260
 QY 7261 TCCCTCCGTTGATACAGGATGCGGATTCGCACTTACCAAGGCTTCCTCTGAGGAGTTCCTCTGAGG 7320
 Db 7261 TCCCTCCGTTGATACAGGATGCGGATTCGCACTTACCAAGGCTTCCTCTGAGGAGTTCCTCTGAGG 7320
 QY 7321 CAGGAGAGAGGAGGAGTTCCTCTGAGGATTCGCACTTACCAAGGCTTCCTCTGAGGAGTTCCTCTGAGG 7380
 Db 7321 CAGGAGAGAGGAGGAGTTCCTCTGAGGATTCGCACTTACCAAGGCTTCCTCTGAGGAGTTCCTCTGAGG 7380
 QY 7381 TCCGCACTAAGACTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7440
 Db 7381 TCCGCACTAAGACTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7440
 QY 7441 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7500

RESULT 15
 ID ABK91430
 AC ABK91430;
 DT 15-NOV-2002 (first entry)
 XX Hepatitis C virus Con 1 isolate DNA mutant 7.
 DE HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
 KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
 KW internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
 XX Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 342..9374
 FT /*tag= a
 FT /product= "HCV polyprotein"
 FT /note= "The polyprotein consists of the Core, E1, E2, P7,
 FT NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
 FT mutation replace(6931,C)
 FT /*tag= b
 FT XX
 PN W0200259321-A2.
 PD 01-AUG-2002.
 XX
 PF 16-JAN-2002; 2002WO-EP000526.
 XX
 PR 23-JAN-2001; 2001US-0263479P.
 XX
 PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
 XX
 PI De Francesco R, Migliaccio G, Paonessa G;
 XX
 DR WPI; 2002-599793/64.
 XX
 PT New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
 PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
 PT ribosome entry site (IRES) region, useful in studying HCV replication and
 PT expression.
 PS
 PS Claim 9; Page; 69pp; English.
 XX
 CC The invention relates to nucleic acid molecules comprising altered HCV
 CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
 CC internal ribosome entry site (IRES) region coding for one or more NS3,
 CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
 CC are detailed in the specification. Also included are (1) an expression
 CC vector comprising a nucleotide sequence coding for the altered nucleic
 CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
 CC recombinant cell human hepatoma cell comprising the altered nucleic acids
 CC ; (3) a recombinant cell produced by introducing into a human hepatoma
 CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
 CC replicon enhanced cell or which containing a functional HCV replicon; (5)
 CC an HCV replicon enhanced cells made in the method; and (6) measuring the
 CC ability of a compound to affect HCV activity. The HCV replicons and HCV
 CC replicon enhanced cells are useful in studying HCV replication and
 CC expression, and HCV and host cell interactions, producing HCV RNA and
 CC proteins, and providing a system for measuring the ability of a compound
 CC to modulate one or more HCV activities e.g. to discover drugs which may
 CC treat HCV mediated diseases such as liver failure, cirrhosis and
 CC hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
 CC mutant of the invention. Note: The present sequence is not shown in the
 CC specification but was created by the indexer using the HCV sequence
 CC appearing as ABK91411 and the information in Claim 9
 XX
 SQ Sequence 9605 BP; 1910 A; 2882 C; 2733 G; 2060 T; 0 U; 0 Other;

Query Match 86.0%; Score 8248.6; DB 6; Length 9605;
 Best Local Similarity 91.3%; Pred. No. 0;
 Matches 8786; Conservative 0; Mismatches 829; Indels 10; Gaps 1;
 QY 1 GCCAGCCCCCTGATGGGGGCGACACTCCACCATGATCACTCCCTGTGAGGAATCTACTG 60
 DB |||||
 QY 1 GCCAGCCCCGATGGGGGCGACACTCCACCATAGATCACTCCCTGTGAGGAATCTACTG 60
 DB |||||
 QY 61 TCTTCAGCAGAAAGCGTCTAGCCATGGCCTTAGTATGAGTGTCTGAGAGCTCCAGGAC 120
 DB |||||
 QY 61 TCTTCAGCAGAAAGCGTCTAGCCATGGCCTTAGTATGAGTGTCTGAGAGCTCCAGGAC 120
 DB |||||
 QY 121 CCCCCCTCCCGGAGAGCCATAGTGTCTGCGGAAACCGGTAGATACACCGGAATGCCAG 180
 DB |||||
 QY 121 CCCCCCTCCCGGAGAGCCATAGTGTCTGCGGAAACCGGTAGATACACCGGAATGCCAG 180
 DB |||||
 QY 181 GACGACCGGGTCTTCTTGGATCAACCGCTCAATGCCGTGAGATTTGGGGTGCCTCC 240
 DB |||||
 QY 181 GACGACCGGGTCTTCTTGGATCAACCGCTCAATGCCGTGAGATTTGGGGTGCCTCC 240
 DB |||||
 QY 241 GCGAGACTGCTAGCCGAGTAGTGTGGGTGCGGAAAGCGCTTGTGCTACTGCTCATAGG 300
 DB |||||
 QY 241 GCGAGACTGCTAGCCGAGTAGTGTGGGTGCGGAAAGCGCTTGTGCTACTGCTCATAGG 300
 DB |||||
 QY 301 GTGCTTGGAGTGCCCCCGGAGGTCTCGTAGACCGTGCACCATGAGCAGATCCTAAAC 360
 DB |||||
 QY 301 GTGCTTGGAGTGCCCCCGGAGGTCTCGTAGACCGTGCACCATGAGCAGATCCTAAAC 360
 DB |||||
 QY 361 CTCAAAGAAAAACCAAAACGTAAACACCAACCGCGGCCACAGGACGTCAGTTCCCGGGG 420
 DB |||||
 QY 361 CTCAAAGAAAAACCAAAACGTAAACACCAACCGCGGCCACAGGACGTCAGTTCCCGGGG 420
 DB |||||
 QY 421 GTGGTCAGATGTTTGGTGGAGTTTACCTTGTCCCGGAGGGGCCCGAGGTGGGTGTGC 480
 DB |||||
 QY 421 GTGGTCAGATGTTTGGTGGAGTTTACCTTGTCCCGGAGGGGCCCGAGGTGGGTGTGC 480
 DB |||||
 QY 481 GCGGACTAGGAAGCTTCCGAGCGGTGCAACCTCGTGGAGGGGCAACCTATCCCAA 540
 DB |||||
 QY 481 GCGGACTAGGAAGCTTCCGAGCGGTGCAACCTCGTGGAGGGGCAACCTATCCCAA 540
 DB |||||
 QY 541 AGGCTGCGGACCCCGAGGCGAGGGCTTGGGCTCAGCCCGGGTACCCCTTSGCCCTCTATG 600
 DB |||||
 QY 541 AGGCTGCGGACCCCGAGGCGAGGGCTTGGGCTCAGCCCGGGTACCCCTTSGCCCTCTATG 600
 DB |||||
 QY 601 GCAATGAGGGCTGGGGTGGGAGAGTGGCTTGTGACCCCGGGTCCCGGCTAGTT 660
 DB |||||
 QY 601 GCAATGAGGGCTGGGGTGGGAGAGTGGCTTGTGACCCCGGGTCCCGGCTAGTT 660
 DB |||||
 QY 661 GGGGCCCCACGACCCCCCGGTAGGTGCGCTAACTTGGGTAAAGTCAATCGATACCCCTTA 720
 DB |||||
 QY 661 GGGGCCCCACGACCCCCCGGTAGGTGCGCTAACTTGGGTAAAGTCAATCGATACCCCTTA 720
 DB |||||
 QY 721 CATGGCGTTCGCGATCTCATGGGTACATTCGGCTCGTGGGCGCCCTTAGGGGGCG 780
 DB |||||
 QY 721 CATGGCGTTCGCGATCTCATGGGTACATTCGGCTCGTGGGCGCCCTTAGGGGGCG 780
 DB |||||
 QY 781 CTGCCAGGCGCTTGGCACACCGTGTCCGGGTTCGGAGGCGGTGAGTATGCACAG 840
 DB |||||
 QY 781 CTGCCAGGCGCTTGGCACACCGTGTCCGGGTTCGGAGGCGGTGAGTATGCACAG 840
 DB |||||
 QY 841 GGAATCTGCCCGGTGTCTTCTCTATCTTCTTCTTGGCTCTGTGTCTCTGTGTTGACCA 900
 DB |||||
 QY 841 GGAATCTGCCCGGTGTCTTCTCTATCTTCTTCTTGGCTCTGTGTCTCTGTGTTGACCA 900
 DB |||||
 QY 901 TCCAGGCTTCGCTATGAGTGGCAACGTTGCCGGATATACCACTGTACGAACTGACT 960
 DB |||||
 QY 901 TCCAGGCTTCGCTATGAGTGGCAACGTTGCCGGATATACCACTGTACGAACTGACT 960
 DB |||||
 QY 961 GCTCCAACTCAGCATGTGTATGAGGAGCGGAGCTGTATCATGCACTCTCCGGGTGCG 1020
 DB |||||
 QY 961 GCTCCAACTCAGCATGTGTATGAGGAGCGGAGCTGTATCATGCACTCTCCGGGTGCG 1020
 DB |||||
 QY 1021 TGGCCTGTGTTAGGAGGGTAAACAGTCCCGTCTGCTGGTAGCGCTCACTCCCGGCTCG 1080

Db 1021 TCCCTGGGTTTCGGAGAACAACTCTCCGCTGCTGGTAGCGCTCACTCCACGCTCG 1080
QY 1081 CGGCCAGGAATCCAGCGTCCCACTACGACAAATAGAGCGACGTCGACTGCTCGTGG 1140
Db 1081 CGGCCAGGAACGCTAGCGTCCCACTACGACAAATAGAGCGACGTCGACTGCTCGTGG 1140
QY 1141 GGAAGGCTGCTTCTGCTCCGCTATGATGAGTGGGGATCTCTCGCGATCTATTTTCCTCG 1200
Db 1141 GGGCGCTGCTCTCTGCTCCGCTATGATGAGTGGGGATCTCTCGCGATCTGTTTTCTCG 1200
QY 1201 TCTCCAGCTGTACCTCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 1260
Db 1201 TCGCCAGCTGTACCTCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 1260
QY 1261 CAATCTATCCCGCCATGATCAGGTCAACGCTCCGCTCCGCTCCGCTCCGCTCCGCT 1320
Db 1261 CAATATATCCCGCCATGATCAGGTCAACGCTCCGCTCCGCTCCGCTCCGCTCCGCT 1320
QY 1321 CACCTACAGACCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
Db 1321 CACCTACAGACCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
QY 1381 TGGTCGGGGGCGCCACCTGGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Db 1381 TGGTCGGGGGCGCCACCTGGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
QY 1441 ACTGGGCTAAGGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
Db 1441 ACTGGGCTAAGGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
QY 1501 CGACGGGAGGTCGGCGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
Db 1501 TGAAGGGGGGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
QY 1561 CGTCTCAGAAATCCAGCTTGTGATACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
Db 1561 CATCCAGAAATCCAGCTTGTGATACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
QY 1621 TAAATGCAATGACTCCCTCAAACTGGGTTCTTGGCGGCTGCTGCTGCTGCTGCTGCTG 1680
Db 1621 TGAATGCAATGACTCCCTCAAACTGGGTTCTTGGCGGCTGCTGCTGCTGCTGCTGCTG 1680
QY 1681 TCACTGCTCGGTCGGCGGAGCGCATGGCGAGCTGGCGGAGCTGGCGGAGCTGGCGGAG 1740
Db 1681 TCACTGCTCGGTCGGCGGAGCGCATGGCGAGCTGGCGGAGCTGGCGGAGCTGGCGGAG 1740
QY 1741 AGGGTGGGGGCGGATACCTATACCTAAGCTTACAGCTCGGATCAGAGGCTGCTGCTGCT 1800
Db 1741 AGGGTGGGGGCGGATACCTATACCTAAGCTTACAGCTCGGATCAGAGGCTGCTGCTGCT 1800
QY 1801 GGCATTAAGCGGCTGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
Db 1801 GGCATTAAGCGGCTGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
QY 1861 ATTGTTTACCCCAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
Db 1861 ACTGCTTACCCCAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
QY 1921 ATAGCTGGGGGAGATGAGACAGCTGATGCTTCTCAACAAACAGCGCTCGGCGCAAG 1980
Db 1921 ACAGTTGGGGGAGATGAGACAGCTGATGCTTCTTAAACAAACAGCGCGCGCGCAAG 1980
QY 1981 GCACTGCTTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
Db 1981 GCACTGCTTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
QY 2041 CCCGCTGATACATCGGGGGTTCGCTAACCCGCTGATGATGATGATGATGATGATGATGATG 2100
Db 2041 CCCGCTGATACATCGGGGGTTCGCTAACCCGCTGATGATGATGATGATGATGATGATGATG 2100
QY 2101 GGAAGCACCCGAGCTTACTTACACAAATGTGCTCGGGGCTGCTGCTGCTGCTGCTGCTG 2160

Db 2101 GGAAGCACCCGAGCTTACTTACACAAATGTGCTCGGGGCTGCTGCTGCTGCTGCTGCTG 2160
QY 2161 GCCTAGTAGACTACCCATACAGGCTTTGGCAGCTTACCCCTGACCTCTCAATTTTTCCTCT 2220
Db 2161 GCCTAGTAGACTACCCATACAGGCTTTGGCAGCTTACCCCTGACCTCTCAATTTTTCCTCT 2220
QY 2221 TTAAGGTTAGGATATGTGGGGGCGCTGGAGCAGAGCTCAATGCGGCTGCAATTTTGA 2280
Db 2221 TCAAGGTTAGGATATGTGGGGGCGCTGGAGCAGAGCTCAATGCGGCTGCAATTTTGA 2280
QY 2281 CTCAGAGAGAGCGCTGTAACTTGGAGGACAGGATAGGTGAGAACTCAGCCCGCTGCTG 2340
Db 2281 CTCAGAGAGAGCGCTGTAACTTGGAGGACAGGATAGGTGAGAACTCAGCCCGCTGCTG 2340
QY 2341 TGTCTACACAGAGTGGCAGATCTGCCCTGCTGCTTTCACACCTACCCGCTTTATCCA 2400
Db 2341 TGTCTACACAGAGTGGCAGATCTGCCCTGCTGCTTTCACACCTACCCGCTTTATCCA 2400
QY 2401 CTGCTTATCCATCTCCATCAGAACATCGTGGACGTCGAATACCTGTACGGTGTAGGGT 2460
Db 2401 CTGCTTATCCATCTCCATCAGAACATCGTGGACGTCGAATACCTGTACGGTGTAGGGT 2460
QY 2461 CAGCGTTTGTCTCTTTCGAAATCAATGGGAGTACATCTGCTGCTGCTGCTGCTGCTGCTG 2520
Db 2461 CAGCGTTTGTCTCTTTCGAAATCAATGGGAGTACATCTGCTGCTGCTGCTGCTGCTGCTG 2520
QY 2521 CAGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
Db 2521 CAGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
QY 2581 CCTTAGAGAACTTGGTGTCTCTCAATCGGCGCTCCGTGGCGGAGCGCATGCTGCTGCT 2640
Db 2581 CCTTAGAGAACTTGGTGTCTCTCAATCGGCGCTCCGTGGCGGAGCGCATGCTGCTGCT 2640
QY 2641 CCTTCTTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2700
Db 2641 CCTTCTTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2700
QY 2701 CGTATGCTTTTTATGGGCTATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760
Db 2701 CATATGCTTCTAGCGGCTATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760
QY 2761 CTATGCTTCTAGCGGCTATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820
Db 2761 CATATGCTTCTAGCGGCTATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820
QY 2821 TATCTTGTAGCTTGTACCATATACAAAGTGTCTCTAGCTGCTGCTGCTGCTGCTGCTG 2880
Db 2821 TATCTTGTAGCTTGTACCATATACAAAGTGTCTCTAGCTGCTGCTGCTGCTGCTGCTG 2880
QY 2881 AATATCTTATACAGAGCGGCGCAATGCAAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2940
Db 2881 AATATCTTATACAGAGCGGCGCAATGCAAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2940
QY 2941 GGGAGGCGCGGAGTGCATCTCTCTCAAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3000
Db 2941 GGGAGGCGCGGAGTGCATCTCTCTCAAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3000
QY 3001 ACATCAACCAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3060
Db 3001 CCATACCAAAATCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3060
QY 3061 CGAGAGTGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3120
Db 3061 CCAGAGTGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3120
QY 3121 AATGCGCGGGGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3180
Db 3121 AGTTGCTGGGGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3180
QY 3181 CGTACCTTATACCATCTTACCCACTGCGGAGCTGGGCGCCACCGCGGCTTACGAGAC 3240
Db 3181 CGTACCTTATACCATCTTACCCACTGCGGAGCTGGGCGCCACCGCGGCTTACGAGAC 3240

QY 3241 TTGCGGTGCGGTAGAGCCCGTGTCTTCTCCGCCATGAGACCAAGGTCAACCTGGG 3300
DB |||||
3241 TTGCGGTGCGGTAGAGCCCGTGTCTTCTCTGATGAGACCAAGGTATCACCTGGG 3300
QY 3301 GAGCAGACCCGCTGCTGCTGGGACATCATCTTGGGTCTACCGGTCTCGGCCGCAAGG 3360
DB |||||
3301 GGGCAGACCCGCGGTGTGGGACATCATCTTGGGCCCTGCGCCGTCCTCGGCCGCAAGG 3360
QY 3361 GGAAGAGATATTTTGGGACCGGTGTAGTCTCGAAGGCAAGGTCGCGACTCTCTTG 3420
DB |||||
3361 GGAGGAGATACATCTGGGACCGGCAGACAGCTTGAAGGGCAGGGTGGCACTCTCTG 3420
QY 3421 CGCCCATCAGCGCTACTCCAAACAAACCGGGGGGTACTTGGTGCATCATCATAGCC 3480
DB |||||
3421 CGCCATATACCGCTACTCCCAACACAGCGCGAGGCTACTTGGCTGCATCATCATAGCC 3480
QY 3481 TCACAGGCGGACAGAACACAGGTGCGAAGGGAGGTCAAGTGGTCTACCGCAACAC 3540
DB |||||
3481 TCACAGGCGGACAGAACACAGGTGCGAAGGGAGGTCAAGTGGTCTCCACCGCAACAC 3540
QY 3541 AATCTTCTCGCGACCTCGCATCAAGCGGTGTGTGGACTGTCTACCATGGCGTGGCT 3600
DB |||||
3541 AATCTTCTCGCGACCTCGCATCAAGCGGTGTGTGGACTGTCTCATCTGTGGCGCT 3600
QY 3601 CGAAGACCTTAGCGGTCCAAAGGTCAATCAACCCAAATGACCAATGTAGACCTTGG 3660
DB |||||
3601 CAAAGACCTTTCGCGGCCCAAGGGCCCAATCACCCAAATGTACCAATGTGGACCAAG 3660
QY 3661 ACCTGCTGCGCTGCGAGGCGCCCGCGGCGCTCCATGACACCATGACACTGTGCA 3720
DB |||||
3661 ACCTGCTGCGCTGCGAGGCGCCCGCGGCGCTTCTTGACACCATGACACTGTGCA 3720
QY 3721 GCTCGGACCTTACTTGGTACAGACATGTGTATCTCCGCTGCGCGCGCGAGGCG 3780
DB |||||
3721 GCTCGGACCTTACTTGGTACAGGACATGCGATGTCAATTCOGGTGCGCGCGCGGCG 3780
QY 3781 ACAGCAGGGAAGTCTACTCTCCGAGCGCGCTCTCTTAACCTGAAAGAGCTCTCTGG 3840
DB |||||
3781 ACAGCAGGGAAGTCTACTCTCCGAGCGCGCTCTCTTAACCTGAAAGAGCTCTCTGG 3840
QY 3841 GTCATCTGCTTCCCTTCGCGGACGCTCTGCGGCGCTCTCCGCGCTGCTGTGTCACCC 3900
DB |||||
3841 GTCATCTGCTTCCCTTCGCGGACGCTGTGGGCACTCTTTCGGGCTGCGGTGTGCAACC 3900
QY 3901 GGGGGTTCGGAAGCGGTGGAATTCATACCGTTGAGTCTATGAAACTACCATGCGGT 3960
DB |||||
3901 GAGGGGTTCGGAAGCGGTGGAATTTGTACCGGTGAGTCTATGGAACCACTATGCGGT 3960
QY 3961 CTCGGTCTTTCACAGCACTCAACCCCGCGCTGTACCGAGACATTCGAAGTGGAC 4020
DB |||||
3961 CCCCGGTCTTTCAGGACACTCGTCCCTTCGCGCGGTACCGCAGACATTCGAAGTGGCC 4020
QY 4021 ATCTCACGCTCTACTGCGCAGCGCAAGACCAAGTTCGCGCTGCTGTATGAGGCC 4080
DB |||||
4021 ATCTCACGCTCTACTGAGCGCAAGACACTTAAAGTTCGCGCTGCTGTATGAGGCC 4080
QY 4081 AAGGTACAGGTGTCTGTGTAACCGGTTCGTTGCGCGCACTTAAAGTTCGCGGT 4140
DB |||||
4081 AAGGTATAGGTGTCTGTGTAACCGGTTCGTTGCGCGCACTTAAAGTTCGCGGT 4140
QY 4141 ATATGTCGAAGGCACAGGTATCGACCTTAACATCAGAACTGGGTAGGACCATACCA 4200
DB |||||
4141 ATATGTCGAAGGCACATGGTATCGACCTTAACATCAGAACTGGGTAGGACCATACCA 4200
QY 4201 CGGCGCTTCCATACGACTTCCACTATGGCAAGTTCCTTCGCGAGCGGTGCTTCTG 4260
DB |||||
4201 CGGCTGCCCATCAGTACTTCCACTATGGCAAGTTCCTTCGCGAGCGGTGCTTCTG 4260
QY 4261 GGGGCGCTATGACATCATATATGTGATGAGTGCACCTCAACTGACTCGACTACCATCT 4320
DB |||||
4261 GGGGCGCTATGACATCATATATGTGATGAGTGCACCTCAACTGACTCGACTACCATCT 4320

QY 4321 TGGCATCGCACAGTCTCTGACCAAGCGGAGACGGCTGGAGCGCGCTCTCTGCTGCG 4380
DB |||||
4321 TGGCATCGCACAGTCTCTGACCAAGCGGAGACGGCTGGAGCGCGCTCTCTGCTGCTGCG 4380
QY 4381 CCACCGCTACAGCTCCGGATCGGTACCGTGCACACCCCAATATCGAGGAATAGGCC 4440
DB |||||
4381 CCACCGCTACCGCTCCGGATCGGTACCGTGCACACCCCAATATCGAGGAATAGGCC 4440
QY 4441 TGTCCAAATATGGAGATCCCTTCTATGGCAAGCCATCCCATTCGAGGCCATCAAGG 4500
DB |||||
4441 TGTCCAGCACTGGAGATCCCTTCTATGGCAAGCCATCCCATTCGAGGCCATCAAGG 4500
QY 4501 GGGGAGGCACTCATTTTCTGSCANTTCCAAAGAAATGTGACGAGTCCCGCAAGC 4560
DB |||||
4501 GGGGAGGCACTCATTTTCTGSCANTTCCAAAGAAATGTGATGAGTCCCGCAAGC 4560
QY 4561 TGACAGGCTCGGACTGACCTGTAGCATATTAACGGGCGCTTATGTGCTCCCTCATAC 4620
DB |||||
4561 TGTCCGGCTCGGACTCAATGCTGTAGCATATTAACGGGCGCTTATGTGCTCCCTCATAC 4620
QY 4621 CGCCTATCGGAGACGCTGCTGCTGSCAAACAGACGCTCTAATGACGGGTTTCAACGGCG 4680
DB |||||
4621 CAACCTAGCGGAGACGCTCATTTCTGTAGCAACGAGCGCTCTAATGACGGGTTTCAACGGCG 4680
QY 4681 ATTTGACTCATGTGATGACTGCAATATCATGTGTCAACGACAGTCTGACTTCAGCTTGG 4740
DB |||||
4681 ATTTGACTCATGTGATGACTGCAATATCATGTGTCAACGACAGTCTGACTTCAGCTTGG 4740
QY 4741 ATCCACCTTCAACATTCAGACGAGCGCGTCCCAAGACGGGTGCGGCTCGCAAC 4800
DB |||||
4741 ACCGACCTTCAACATTCAGACGAGCGCGTCCCAAGACGGGTGCGGCTCGCAAC 4800
QY 4801 GCGAGGTAGAACTGCGAGGCTAGGAGTGCATCTACAGGTTTGTGACTCCAGAGAAC 4860
DB |||||
4801 GCGAGGACGACTGTGTAGGCGAGGATGSCATTTACAGGTTTGTGACTCCAGAGAAC 4860
QY 4861 GGCCTTCGGGCAATGTTGATTTCTTGGTCTGTGTGAGTGTATGACGCGGCTTGTGCTT 4920
DB |||||
4861 GGCCTTCGGGCAATGTTGATTTCTTGGTCTGTGTGAGTGTATGACGCGGCTTGTGCTT 4920
QY 4921 GGTATGAGTCAACCGCGCTGAGACCTCGGTAGGTTGCGGCTTCACTAATACACAG 4980
DB |||||
4921 GGTATGAGTCAACCGCGCTGAGACCTCGGTAGGTTGCGGCTTCACTAATACACAG 4980
QY 4981 GGTTCGCGCTGCGCAGACCATCTGGAGTTCGCGAGAGGCTCTTACAGGCTCAACC 5040
DB |||||
4981 GGTTCGCGCTGCGCAGACCATCTGGAGTTCGCGAGAGGCTCTTACAGGCTCAACC 5040
QY 5041 ACATAGATCCCACTTCTGTCGAGACTAAACAGGCGAGAGCAACTTCCCTTACCTGG 5100
DB |||||
5041 ACATAGATCCCACTTCTGTCGAGACTAAACAGGCGAGAGCAACTTCCCTTACCTGG 5100
QY 5101 TGGCATATCAAGCTTACAGTGTGCGCAGGCTCAAGCTCCACCTTCCATCTGGGACCAAA 5160
DB |||||
5101 TAGCATACAGGCTACGTTGTGCGCAGGCTCAGGCTCCACCTCCATCTGTGGAGCAAA 5160
QY 5161 TGTGGAAGTGTCTCATACGCTAAAGCTACGCTGACAGGCGCAACGCGCTGTGTATA 5220
DB |||||
5161 TGTGGAAGTGTCTCATACGCTAAAGCTACGCTGACAGGCGCAACGCGCTGTGTATA 5220
QY 5221 GGCTAGGAGCGCTCCAAATGAGGTCTCTTCCACACACCCCATTAACATACTATG 5280
DB |||||
5221 GGCTAGGAGCGCTCCAAATGAGGTCTCTTCCACACACCCCATTAACATACTATG 5280
QY 5281 CATGATCTCGCTGACTGCTGAGGTCTGCTAGCACTGAGTCTGCTAGGCGGAGTCC 5340
DB |||||
5281 CATGATCTCGCTGACTGCTGAGGTCTGCTAGCACTGAGTCTGCTAGGCGGAGTCC 5340
QY 5341 TTGAGCTTTGGCGGCATCTATGCTGACGAGAGTGTGGTCAATTTSTGGGAGGATCA 5400
DB |||||
5341 TAGCAGCTTGGCGCGGTATTTGCTGACACAGGACGCTGCTCATTTGTGGGAGGATCA 5400
QY 5401 TCTTGTCCGGAGACGAGCTGCTGTTCCCGACAGGAGTCTCTTACAGAGGTTCGATG 5460

QY 7621 GCGCCCTGATCAGCCATGCGCTCGGAGGAAAGTAAGTGCCTCCATCAACCCCTTGAGCA 7680
 Db 7621 GCGCCCTGATCAGCCATGCGCTCGGAGGAAAGTAAGTGCCTCCATCAACCCCTTGAGCA 7680
 QY 7681 ACTCTTTGCTGCTCACCACAAATAGTCTACGGCACAACATCCCGAGGCGAAGCTCC 7740
 Db 7681 ACTCTTTGCTGCTCACCACAAATAGTCTACGGCACAACATCCCGAGGCGAAGCTCC 7740
 QY 7741 GGCAGAGAAGAGTCACTTTGACAGATTGCAAGTCTCGTGAATGATCATATACCGGAGAGTAC 7800
 Db 7741 GGCAGAGAAGAGTCACTTTGACAGATTGCAAGTCTCGTGAATGATCATATACCGGAGAGTAC 7800
 QY 7801 TCAAGGAGATCAAGGCGAAGGCTCCACAGTTAAGGCTAAGCTTCTATCTATAGAGAGG 7860
 Db 7801 TCAAGGAGATCAAGGCGAAGGCTCCACAGTTAAGGCTAAGCTTCTATCTATAGAGAGG 7860
 QY 7861 CTTGCAAGCTGAGCCGCCCACTTCGCGCCAGATCTAAATTTGGCTATGGGGCAAGAGCG 7920
 Db 7861 CTTGCAAGCTGAGCCGCCCACTTCGCGCCAGATCTAAATTTGGCTATGGGGCAAGAGCG 7920
 QY 7921 TCCGGAACCTATCAGCAGGCGGCTTAAACACATCCGCTCCGCTGGGAGGACTTGTGG 7980
 Db 7921 TCCGGAACCTATCAGCAGGCGGCTTAAACACATCCGCTCCGCTGGGAGGACTTGTGG 7980
 QY 7981 AAGACACTGAAACACCAATGACACCAATCATGSCAATAAGTGAAGTTTCTGGGTCC 8040
 Db 7981 AAGACACTGAAACACCAATGACACCAATCATGSCAATAAGTGAAGTTTCTGGGTCC 8040
 QY 8041 AACCAGAGAGGAGCGCCGACAGCTGCGCTTATCGTATCCAGACAGCTGGGAGTTC 8100
 Db 8041 AACCAGAGAGGAGCGCCGACAGCTGCGCTTATCGTATCCAGACAGCTGGGAGTTC 8100
 QY 8101 GTGTATGCGAGAGATGCGGCTTTACAGCTGTGCTCCACCTTCTCAGGCGCTGATGG 8160
 Db 8101 GTGTATGCGAGAGATGCGGCTTTACAGCTGTGCTCCACCTTCTCAGGCGCTGATGG 8160
 QY 8161 GCTCTCATACCGATTTCATCTCCGAGAGGCTGAGTTCCTGCTGCAATACCT 8220
 Db 8161 GCTCTCATACCGATTTCATCTCCGAGAGGCTGAGTTCCTGCTGCAATACCT 8220
 QY 8221 GGAATTCAGAGAAATGCGCTATGGCTTCTCATATGACACCGCTGTTTGTGCTAACGG 8280
 Db 8221 GGAATTCAGAGAAATGCGCTATGGCTTCTCATATGACACCGCTGTTTGTGCTAACGG 8280
 QY 8281 TCACTGAGAGTGAATTCGTTGAGAGTCAATTTACCAATGTTGACTTGGCCCGG 8340
 Db 8281 TCACTGAGAGTGAATTCGTTGAGAGTCAATTTACCAATGTTGACTTGGCCCGG 8340
 QY 8341 AGCCAGACAGGCGCTAAGTGTCTCAGAGCGGCTTTACATCGGGGTCCTCCCTGACTA 8400
 Db 8341 AGCCAGACAGGCGCTAAGTGTCTCAGAGCGGCTTTACATCGGGGTCCTCCCTGACTA 8400
 QY 8401 ACTCAAAAGGCGAGAACTCGGCTATCGCGGTCGCGGCTGAGTGCCTGACGACTA 8460
 Db 8401 ACTCAAAAGGCGAGAACTCGGCTATCGCGGTCGCGGCTGAGTGCCTGACGACTA 8460
 QY 8461 GCTGGGTAAATACCTTCAATGTTACTTGAAGGCCACTGACGCTGTGAGCTGCAAAAGC 8520
 Db 8461 GCTGGGTAAATACCTTCAATGTTACTTGAAGGCCACTGACGCTGTGAGCTGCAAAAGC 8520
 QY 8521 TCCAGACTGACAGATGCTCGTGAACGGAGAGAGACCTTCTGTTAATCTGTGAAGCGGG 8580
 Db 8521 TCCAGACTGACAGATGCTCGTGAACGGAGAGAGACCTTCTGTTAATCTGTGAAGCGGG 8580
 QY 8581 GAACCCAGGAGATGCGGCGCCCTACGAGCTTCAAGGCTTCAAGGCTATGACTATTCG 8640
 Db 8581 GAACCCAGGAGATGCGGCGCCCTACGAGCTTCAAGGCTTCAAGGCTATGACTATTCG 8640
 QY 8641 CCCCCCCTGGGATCGGCCCAACAGATAAGCCTGAGCTGATAACATCATGTTCT 8700
 Db 8641 CCCCCCCTGGGATCGGCCCAACAGATAAGCCTGAGCTGATAACATCATGTTCT 8700

QY 8701 CCAATGTGTCACTGCGCACAGATGATCTGGCAAAAGGGTATACTACCTCACCCGTGACC 8760
 Db 8701 CCAATGTGTCACTGCGCACAGATGATCTGGCAAAAGGGTATACTATCTCACCCGTGACC 8760
 QY 8761 CACACACCCCTTGTGACGGGTGCGTGGAGACAGCTAGACACACTCCAAATCAACTCTT 8820
 Db 8761 CACACACCCCTTGTGCGGGTGGTGGGAGACAGCTAGACACACTCCAGTCAATCTCT 8820
 QY 8821 GCGTAGGCAATATCATGATGATGCGCCACCTATGGGCAAGGATGATCCCTGATGACTC 8880
 Db 8821 GCGTAGGCAATATCATGATGCGCCACCTATGGGCAAGGATGATCCCTGATGACTC 8880
 QY 8881 ACTTTTCTTCCATCTCTAGCTCAAGAGCAACTTTGAAAAGCCCTGGATGTCAGATCT 8940
 Db 8881 ATTCTCTTCCATCTCTAGCTCAAGAGCAACTTTGAAAAGCCCTGGATGTCAGATCT 8940
 QY 8941 AGGGGCTTGTCTACTTCCATTTAGGACCACTTGAACCTACCTCAGATCATTTGAACCTCATG 9000
 Db 8941 AGGGGCTTGTCTACTTCCATTTAGGACCACTTGAACCTACCTCAGATCATTTGAACCTCATG 9000
 QY 9001 GTCTTAGCGCATTTTACACTCCACAGTTTACTCTCCAGGTGAGATCAATAGGGTGGCTTCA 9060
 Db 9001 GCTTAGCGCATTTTACACTCCATAGTTTACTCTCCAGGTGAGATCAATAGGGTGGCTTCA 9060
 QY 9061 GCGTCAGGAACTTGGGTTACCACTTGGGAACTGGAGACATCGGGCCAGAACTGTCC 9120
 Db 9061 GCGTCAGGAACTTGGGTTACCACTTGGGAACTGGAGACATCGGGCCAGAACTGTCC 9120
 QY 9121 GCGTTAAGTACTGTCCAGGGGGGAGGCGGCCACATTTGGTGGGAGATACCTCTTTAAT 9180
 Db 9121 GCGTTAAGTACTGTCCAGGGGGGAGGCGGCCACATTTGGTGGGAGATACCTCTTTAAT 9180
 QY 9181 GGGCAGTAGGAGCAAGCTTAAACTCACTTCAATCCCGCGCGCTCCAGCTGACTTGT 9240
 Db 9181 GGGCAGTAGGAGCAAGCTTAAACTCACTTCAATCCCGCGCGCTCCAGCTGACTTGT 9240
 QY 9241 CTGGCTGTGTTCTGCTGCTTACAGCGGGGAGACATATATCACAGCTGTCTCGTGCC 9300
 Db 9241 CCAGCTGTGTTCTGCTGCTTACAGCGGGGAGACATATATCACAGCTGTCTCGTGCC 9300
 QY 9301 GACCCCGCTGCTTCCGTTGCTGCTTCTCTAATTTCTGAGGGTAGGCTTTACCTGC 9360
 Db 9301 GACCCCGCTGCTTCCGTTGCTGCTTCTCTAATTTCTGAGGGTAGGCTTTACCTGC 9360
 QY 9361 TCCCAACCCGATGAAACGGGAGCTAAACCACTCCAGGCTTAAAGC-----ATTTC 9410
 Db 9361 TCCCAACCCGATGAAACGGGAGCTAAACCACTCCAGGCTTAAAGC-----ATTTC 9410
 QY 9411 CTGTT 9470
 Db 9411 CTGTT 9470
 QY 9471 TTTTCTTTTCTTTTCCCTTCTTAAAGTGGTGGCTTCAATCTTAGGCTTAGTCAAGGCTAGC 9530
 Db 9471 TTTTCTTTTCTTTTCCCTTCTTAAAGTGGTGGCTTCAATCTTAGGCTTAGTCAAGGCTAGC 9530
 QY 9531 TGTGAAGAGTCCGTGAGCGGATGACTGACAGAGAGTGTGATCTGAGCTCTCTGAGAT 9590
 Db 9531 TGTGAAGAGTCCGTGAGCGGATGACTGACAGAGAGTGTGATCTGAGCTCTCTGAGAT 9590
 QY 9591 CATGT 9595
 Db 9601 CAAGT 9605

Search completed: April 9, 2004, 07:58:44
 Job time : 3251 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	8311.8	86.6	9413	10	US-09-627-688-6	Sequence 6, Appl
2	7885.8	52.2	9275	14	US-10-259-273-39	Sequence 39, Appl
3	6335.8	56.0	9599	12	US-10-189-353-13	Sequence 13, Appl
4	6264	65.3	12980	9	US-09-238-076-5	Sequence 5, Appl
5	6264	65.3	12980	10	US-09-595-931-5	Sequence 5, Appl
6	6264	65.3	12980	10	US-09-517-563-5	Sequence 5, Appl
7	6261.2	65.3	9645	9	US-09-742-659-3	Sequence 3, Appl
8	6254.8	65.2	9645	9	US-09-238-076-1	Sequence 1, Appl
9	6254.8	65.2	9645	10	US-09-595-931-1	Sequence 1, Appl
10	6254.8	65.2	9645	10	US-09-517-563-1	Sequence 1, Appl
11	6149	84.1	10803	9	US-09-747-419-17	Sequence 17, Appl
12	6149	84.1	10803	14	US-10-259-273-17	Sequence 17, Appl
13	6144.4	54.0	9379	9	US-09-916-359-1	Sequence 1, Appl
14	6135.6	63.9	9416	9	US-09-238-076-19	Sequence 19, Appl
15	6135.6	63.9	9416	10	US-09-595-931-19	Sequence 19, Appl

QY	194	TTTCTTGGATCAACCCGCTCAATGCTGGAGATTTTGGCGGTGCCCGCGAGCTGCTAG	255
DB	182	TTTCTTGGATCAACCCGCTCAATGCTGGAGATTTTGGCGGTGCCCGCGAGCTGCTAG	241
QY	254	CCGAGTAGTGTGGGTGCGAAAGGCCTTGTGCTACTGCTCATAGGTTGCTTGGCGAGTG	313
DB	242	CCGAGTAGTGTGGGTGCGAAAGGCCTTGTGCTACTGCTCATAGGTTGCTTGGCGAGTG	301
QY	314	CCCGGGAGTCTCTGTAGACCTGTGACCACTGACCACTGACCACTCAAGAAAAAC	373
DB	302	CCCGGGAGTCTCTGTAGACCTGTGACCACTGACCACTGACCACTCAAGAAAAAC	361
QY	374	CAAAAGTAACCAACACCGCGCCACAGAGACTGACGTTCCGGCGGTGCTCAGATCGT	433
DB	362	CAAAAGTAACCAACACCGCGCCACAGAGACTGACGTTCCGGCGGTGCTCAGATCGT	421
QY	434	TGGTGGAGTTTACCTGTTTCCCGCGAGGGGCCACAGTTTGGTGTGCGCGACATAGGAA	493
DB	422	TGGTGGAGTTTACCTGTTTCCCGCGAGGGGCCACAGTTTGGTGTGCGCGACATAGGAA	481
QY	494	GGCTTCCAGAGCGTTCGCAACCTCTGTGGAAAGCGCAACAATATCCAAAGGTTCGCGACC	553
DB	482	GACTTCCAGAGCGTTCGCAACCTCTGTGGAAAGCGCAACAATATCCAAAGGTTCGCGGCC	541
QY	554	CGAGGCGAGGCGCTGGGCTCAGCCCGGTACCCCTTGGCCCTCTATGCAATGAGGCGCT	613
DB	542	CGAGGTAAGACCTGGGCTCAGCCCGGTACCCCTTGGCCCTCTATGCAATGAGGCGCT	601
QY	614	GGGFTGGCAAGATGGCTCTGTCAACCCCGCGGCTCCCGGCTAGTTGGGCGCCACACGGA	673
DB	602	GGGFTGGCAAGATGGCTCTGTCAACCCCGCGGCTCCCGGCTAGTTGGGCGCCACACGGA	661
QY	674	CCCCGGGTAGGTTCGCTTAATGGGTAAAGTCAATGATACCTTAAATCGCGGCTTCGC	733
DB	662	CCCCGGGTAGGTTCGCTTAATGGGTAAAGTCAATGATACCTTAAATCGCGGCTTCGC	721
QY	734	CGATCTCATGGGGTACATTCGCTCGTCGCGCGCCCTTACGGGCGCTGTCAGAGGCGCT	793
DB	722	CGACTCATGGGGTACATTCGCTCGTCGCGCGCCCTTACGGGCGCTGTCAGAGGCGCT	781
QY	794	GGCAACGCTTCGGGTTCTGGAGGACGGCTGAACCTATGCAACAGGGAATCTGCCCGG	853
DB	782	GGCAACGCTTCGGGTTCTGGAGGACGGCTGAACCTATGCAACAGGGAATCTGCCCGG	841
QY	854	TTGCTCTTCTCATCTTCTCTTGGCTTCGCTGCTCTGTTGACCATCCAGCTTCGCG	913
DB	842	TTGCTCTTCTCATCTTCTCTTGGCTTCGCTGCTCTGTTGACCATCCAGCTTCGCG	901
QY	914	TTATGAAGTGGCAACCTGTCCGGGATATACATGTACAGAACGACTGCTCCAACCTCAAG	973
DB	902	TTACGAGTGGCAACCTGTCCGGGATATACATGTACAGAACGACTGCTCCAACCTCAAG	961
QY	974	CAATGTGATAGGACGGGACGTGATGCAATGCAATCTCCGGGTGCGGCCCTGTGTCA	1033
DB	962	TAATGTGATAGGACGGGACGTGATGCAATGCAATCTCCGGGTGCGGCCCTGTGTCA	1021
QY	1034	GGAGGGTAAACAGCTCCCGTCTCGGTAGCGCTCAGTCCACGCTCCGGGACAGAAATG	1093
DB	1022	GGAGAGTAATTTCTCCGTTGCTGGGTAGCGCTCACTCCACGCTCCGGGACAGAAATG	1081
QY	1094	CAGCGTCCCACATACGACAAATACGACGCGAGCTCGACTTGTCTGCTGGGACGGCTGCTT	1153
DB	1082	CAGCATCCCACACGACAAATACGACGCGAGCTCGACTTGTCTGCTGGGACGGCTGCTT	1141
QY	1154	CTGCTCCGCTATGTACGTGGGGATCTCTGCGGATCTATTTCTCTGCTCTCCAGCTGTT	1213
DB	1142	CTGTTCCGCTATGTACGTGGGGATCTCTGCGGATCTGCTGCTCTCTCCAGCTGTT	1201
QY	1214	CACCTTCTCGCTCCGCGGATAGACAGTACGAGCTGACATGCTCATCTATCCCGG	1273
DB	1202	CACCTTCTCACCTCCGCGGTATGACGGTACAAGATTGAATGCTCAATCTATCCCGG	1261

QY	1274	CCATGTATCAGGTCACCGCATCGGCTTGGGATATGATGAATCACTGGTCACTACAACAGC	1331
DB	1262	CCACGTATCAGGTCACCGCATCGGCTTGGGATATGATGATGAATGGTCACTACAACGGC	1321
QY	1334	CCTAGTGGTGTCGAGTTGCTCCGGATCCACACAGCTGTGTTGAGCATGTTGGCGGGGC	1393
DB	1322	CCTAGTGGTATCGCAGCTACTTCCGGATCCACACAGCGCTGTGACATGTTGCGGGGC	1381
QY	1394	CCACTGGGAGTCTCTGGCGGGCTTGCTACTATTTCATGTTAGGAACTTGGCTAAGT	1453
DB	1382	CCACTGGGCTCTCTGGCGGGCTTGCTACTATTTCATGTTGGGAACTTGGGCTAAGT	1441
QY	1454	TCTGATGTGGCGCTACTCTTTTGGCGGCTGTGACGGGACACCAACACGACGGAGGT	1513
DB	1442	CTTGATTTGTATGTACTCTTTTGGCGGTGACGGGACACCAACCACTGACAGGGGAAG	1501
QY	1514	GGCGGCCACACCCTCCGGCTTCAGTCCCCTTTTCTCATCTGGGGCGTCTCAGAAAT	1573
DB	1502	GGTAGCTTCAGCACCCAGAGCTTCGTGCTCTCAGAGGCCACTCTCAGAAAAT	1561
QY	1574	CCAGCTGTGTAATACCAACGCGAGCTGGCACATCAACAGAGCTCCCTCAANTGCAATGA	1633
DB	1562	CCAACCTGTGAACACCAACGCGAGCTGGCACATCAACAGGACCGCTCTGNATTGCAATGA	1621
QY	1634	CTCCCTCAAACCTGGGTCTTTTGGCGCTGTGTTTAGCACACAAAGTTCAACTCGTCCGG	1693
DB	1622	CTCCCTCAAACCTGGGTCTATTGTCGGCTGTCTACAGAGGCCACTCAAOCGCTCCGG	1681
QY	1694	GTGCCCGGAGGCATGGCAGACTGCCGCCCCCATTTGACTGGTTCCGCCAGGGTGGGGCCC	1753
DB	1682	GTGCCCGAGCGCATGGCTAGCTAGCGGCCCCATTCGATGAGTTCCGCTCAGGGGTGGGGTCC	1741
QY	1754	CATCACCTATAAAGCCTAACAGCTCGGATCAGAGGCCTATTGCTGGCATTTACGCGCC	1813
DB	1742	CATCACTATGATATGCTCAGAGCTCGACACAGAGGCNAATGCTGGCATCTACGCGCC	1801
QY	1814	TCGACCTGTGGTGTCTAACCOCGTGCGAGGTGTGTGFTCCAGTGTATTGTTTCACCCC	1873
DB	1802	TCGACCGTGGGATCGTGCCTGCTCGCAGGTGTGTGTCTCACTGTATTGCTTCACTCC	1861
QY	1874	AAGCCCTGTTGTTGGTGGGACCAACCATCGTTCCGGTGTCCCTACGTATAGCTGGGGGA	1933
DB	1862	GAGCCCTGTTGTATGGGAGCAACCATCGTTTCGGCGCTCCTACGTATAGCTGGGGGA	1921
QY	1934	GAATGACACAGAAGTGTGCTCTCAACACACGCGTCCGCCACAAGGCAACTGTTGCG	1993
DB	1922	GAATGACACAGAAGTGTGCTCTTAGCACACGCGCGCCTCAAGSACAATGCTTTGG	1981
QY	1994	CTGTACATGGATGAATAGTACTGGTTCATTAAGACGTGCGAGGTCCCCCTGTAACT	2053
DB	1982	GTGCACGTGATGAACAGCACTGGGTTCACAAGAAGTGGGGGGCTCCGTGTGAACAT	2041
QY	2054	CGSGGGGTGGTAAACGCACTGTATCTGCCCGAGGACTGTTCCGGAAGCAACCCCA	2113
DB	2042	CGSGGGGTGGCAACAACCTTGTGTGTGCCCCAGGATGCTTCCGGAAGCAACCCCA	2101
QY	2114	GGCTACTTACAAAAATGTGGCTCGSGGGCCCTGGTTGACACCTAGTGTGCTAGTAGACTA	2173
DB	2102	GGCCACTTACAAAGTGTGCTCGSGGGCCCTGGTTGACACCCAGSTGTGATGTTACTA	2161
QY	2174	CCCATCAGGCTTTGGCACTACCCCTGCACTCTCAATTTTTTCCATCTTTAAGGTTAGSAT	2233
DB	2162	CCCATCAGGCTTTGGCACTACCCCTGCACTGTTAACTTTTACGTTTAAAGTCAAGAT	2221
QY	2234	GTATGTGGGGGGTGGAGCACAGGCTCAATGCCGATGCAATTTGATCTCGAGGAGAGCG	2293
DB	2222	GTATGTGGGGGGTGGAGCACAGGCTCAATGCTGATGCAATTTGATCTCGAGGAGAGCG	2281
QY	2294	CTGTAACTTGGAGGACAGGGATAGGTCAAGACTCACGCCGCTGCTGCTGTCTACAAACA	2353
DB	2282	CTGTACTTGGAGGACAGGGATAGGTCAAGACTCACGCCGCTGCTGCTGTCTACAAACA	2341
QY	2354	GTGGCAGATACTCCCTGTGCTTTTCAACCCCTACCGGCTTTATCACTGGTTTGTATCCA	2413

[illegible]

Db	3422	CTATTCCAAACAAACCGCGGCGCTGCTTGCTGTATCATCATCAGCTCAGGTCCGGA	3481
Qy	3494	CAAGAACCAAGCTCGAAGGGGAGGTTCAAGTGGTTTCTACCGCAACAACAATCTTTTCCCTGGC	3553
Db	3482	CAAGAACCAAGTGTATGGGGAGGTTCAAGTGTCTTCCACCGCAACGCAATCTTTCCTGGC	3541
Qy	3554	GACCTGCATCAACGGCGGTGTCTGGACTGTCTTACCATTGGCGCTGGCTCGAAGACCCTTACG	3613
Db	3542	GACCTGCGTCAATGGCGGTGTGTGGACCGTCTACCATGTGTCCCGGCTCGAAGACCCCTGGC	3601
Qy	3614	CGGTCCAAAGGTCCAATCAACCAATGTATACCAATGTATAGACTGGACTCGTGGCTG	3673
Db	3602	CGGCCGAAGGTCCAATCAACCAATGTATACCAATGTATAGACTGGACTCGTGGCTG	3661
Qy	3674	GCAGGCGCCCCCGGGCGCGCTCCATGACACCATGCACTGTGGCACTGCGACACCTTTA	3733
Db	3662	GCGGCGCCCCCGGGCGCGCTCCATGACACCGTGCACCTGCGGCACTGCGACACCTTTA	3721
Qy	3734	CTTGGTCAOGAGACATGCTGATGTCATTCGCGTGCOCGGCGAGGCGACAGCAGGGGAAG	3793
Db	3722	CTTGGTCAOGAGCATGCTGATGTCGTTCCGTTGCGCGCGGGCGACAGCAGGGGGAG	3781
Qy	3794	TCTACTCTCCCGACGGCCGCTCTCTACCTGAAGAAGCTCCTCGGGTGGTTCATTGCTTTG	3853
Db	3782	CCTGCTTTCCCGACGGCCCATCTCTCACTCTGAAGGGTCTCTCGGGTGGACCACTGCTTTG	3841
Qy	3854	CCCTTCGGGCGACGTCGTGGCGGCTCTTCGGGCTGCTGTGCACCCGGGGGTTCGCGAA	3913
Db	3842	CCCTTCGGGCGACGTTGTAGGCATCTTCCGGGCTGCTGTGTGCACCCGGGGGTTGCGAA	3901
Qy	3914	GGCGGTGGACTTCATACCCGTTGAGTCTATGGAACACTTACCATCGGTCCTCGGTTCTCAC	3973
Db	3902	GGCGGTGGACTTCATACCCGTTGAGTCTATGGAACACTTACCATCGGTCCTCGGTTCTCAC	3961
Qy	3974	AGACACTCAACCCCGCGCTGTACCGCAGACATTCCAAGTGGCACATCTGCACGCTCC	4033
Db	3962	AGACAACCTCATCCCTCCGCGCGTACCGCAACATTCCAAGTGGCACATTTACACGCTCC	4021
Qy	4034	TACTGGCAGCGGCAAGACACCAAGTCCCGGCTGGGTATGCAACCAAGGTACAGGT	4093
Db	4022	CATGGCAGCGGCAAGACACCAAGTCCCGGCTGCATATGCAACCAAGGTACAGGT	4081
Qy	4094	GCTCGTCTTCAACCCGTCGCTTCCGCGCACCTTAGGGTTTGGGCGGTATATGTCCAAGGC	4153
Db	4082	GCTCGTCTTCAACCCGTCGCTTCCGCGCACATTTGGAGCGTATATGTCCAAGGC	4141
Qy	4154	ACACGGTATCGACCTTAAGATCAGAACTGGGGTAAGAACCATTTACACCGGCGCTCCAT	4213
Db	4142	ACAATGGCATCGACCTTAACATCAGAACTGGGGTAAGAACCATTTACACCGGCGCGCCAT	4201
Qy	4214	TACGTACTCCACCTATGGCAAGTTCTTTCGCACTGGTGTCTTGGGGCGCTTATGA	4273
Db	4202	CAGTACTCCACCTATTTGCAAGTTCTTTCGCACTGGTGTGTCTTGGGGCGCTTATGA	4261
Qy	4274	CATCATAATATGTATGAGTGCACTCAACTGACTCGACTTGGGCATCGGCAC	4333
Db	4262	CATCATAATATGTATGATGAATGCCACTCAACTGACTCGACTTGGGCATCGGCAC	4321
Qy	4334	AGTCTGGACCAAGCGGAGACGCGTGGAGCGCGGCTGTGCTGCTGCACACCGCTACCC	4393
Db	4322	AGTCTGGATCAGGCAAGACGCGTGGAGCGCGGCTGTGCTGCTGCCACCGCACGCC	4381
Qy	4394	TCCGGGATCGGTTACCGTGCACACCCCAATATCGAGGAAATAGGCTGTCCAACAATGG	4453
Db	4382	TCCGGGATCGATCACCCTGTGCCACACCCCAACATCGAGGAAGTGCGCCCTGTCCACACTGG	4441
Qy	4454	AGAGATCCCTTCTATGGCAAGCCATCCCATTTGAGGGCCATCAAGGGGGGAGGCACTCT	4513
Db	4442	AGAGATTCCTTCTATGGCAAAAGCAATCCCATTTAGGGCCATCAAGGGGGGAGGCACTCT	4501
Qy	4514	CATTTCTGCACTTCCAGAGAAATGTACAGCTGCGCGGAAAGTGTACAGGCTCGG	4573
Db	4502	CATCTTCTGCACTTCCAGAAAGTGTGTGAGCTGCGCGAAAGTGTACAGGCTCGG	4561

4574 ACTGAAAGCTGTAGCATATTACGGGCGCTTGATGTGTCGGTCAATACCGCTATACGAGA 4633
 4562 ACTCAATGCTGTAGCGTATTACGGGGTTCGGATGTGTCGGTATACCGACTAGCGGAGA 4621
 4634 CGTCGTGTGTGGCAACAGACGCTCTAATGACGGGTTCACCGGCGATTTTGAATCACT 4693
 4622 CGTCGTGTGTGGCAACAGACGCTCTAATGACGGGTTCACCGGCGATTTTGAATCACT 4681
 4694 GATCGACTGCAATACATGTGTGACACAGACAGTGCAGCTTCAGCTTGGATTCOACCTTAC 4753
 4682 GATCGACTGCAATACATGTGTGACACAGACAGTGCAGCTTCAGCTTGGATTCOACCTTAC 4741
 4754 CATTCAGACACACACGCTGCCCCAAGACGGGTGTCGGCTTCGCAACGGCGAGTAGAAC 4813
 4742 CATTCAGACACACACGCTGCCCCAAGACGGGTGTCGGCTTCGCAACGGCGAGTAGAAC 4801
 4814 TGGCAGGGGTAGAGTGGCATCTACAGGTTTGTGACTCCAGGAGAACCGGCCCTCCGGCAT 4873
 4802 TGGCAGGGGTAGAGTGGCATCTACAGGTTTGTGACTCCAGGAGAACCGGCCCTCCGGCAT 4861
 4874 GTTCGATTCCTCGCTCCTGTGTGAGTGTGCTATGACGGGGCTGTGCTGTGTATGAGCTCAC 4933
 4862 GTTCGATTCCTCGCTCCTGTGTGAGTGTGCTATGACGGGGCTGTGCTGTGTATGAGCTCAC 4921
 4934 GCCCGCTTGAGACCTCGGTTAGGTTGGGGCTTACCTAAATACACCGGGGTGCCCCCTCTG 4993
 4922 GCCCGCTTGAGACCTCGGTTAGGTTGGGGCTTACCTAAATACACCGGGGTGCCCCCTCTG 4981
 4994 CAGGACCAATCTGGAGTCTGGAGAGCGTCTTACAGGGCTCACCCACATAGATGCCCA 5041
 5054 CTTCTGTGCCACATACAGGACGAGGACACTTTCCTTACTCTGGTGCATATCAAGC 5113
 5042 CTTCTGTGCCACATACAGGACGAGGACACTTTCCTTACTCTGGTGCATATCAAGC 5101
 5114 TACAGTGTGGCGCAGGCTCAAGCTCCACCTCAATCGTGGGACCAATGTGGAGTGTCT 5173
 5102 CACAGTGTGGCGCAGGCTCAAGCTCCACCTCAATCGTGGGACCAATGTGGAGTGTCT 5161
 5174 CATACGGTGAACCTTACTGACGAGGCGCAACACCCCTGCTGTATGAGTGCAGGCGGT 5233
 5162 CATACGGTGAACCTTACTGACGAGGCGCAACACCCCTGCTGTATGAGTGCAGGCGGT 5221
 5234 CCAAAATGAGTGCATCTCTACACACCCCAATACTAAATATACATGATGATGATGATG 5293
 5222 TCAAAATGAGTGCATCTCTACACACCCCAATACTAAATATACATGATGATGATGATG 5281
 5294 TGACCTGGAGGTCTCTACATGACACCTGGGTGCTGTGGTAGGCGGAGTCTTGCAGCTTGGC 5353
 5282 TGACCTGGAGGTCTCTACATGACACCTGGGTGCTGTGGTAGGCGGAGTCTTGCAGCTTGGC 5341
 5354 GCGATCTGCTCAACGACGAGGATGTGTCATGTGGGAGGATCATCTTGTCCGGGAA 5413
 5342 GCGATCTGCTCAACGACGAGGATGTGTCATGTGGGAGGATCATCTTGTCCGGGAG 5401
 5414 GCGAGTGTCTGTTCCCGACAGGGAAGTCTCTTACACAGGATGTCATGAGATGGAAGATG 5473
 5402 GCGAGTGTCTGTTCCCGACAGGGAAGTCTCTTACACAGGATGTCATGAGATGGAAGATG 5461
 5474 TGCTCACAATCTCTTACATGAGCAGGGAATTCAGCTCCCGGAGCAATTCAGGAGGAA 5533
 5462 TGCTCACAATCTCTTACATGAGCAGGGAATTCAGCTCCCGGAGCAATTCAGGAGGAA 5521
 5534 GCGCTCGGCTGTGTGCAACGGCCCAACAGGAGGCTGTGCTCCCTGCTGGA 5593
 5522 GCGCTCGGCTGTGTGCAACGGCCCAACAGGAGGCTGTGCTCCCTGCTGGA 5581
 5594 GTCCAGTGGCGAGCCCTTGAAGCTTCTGGGGAGGACATGATGGAATTTTCACTACAGCGG 5653
 5582 GTCCAGTGGCGAGCCCTTGAAGCTTCTGGGGAGGACATGATGGAATTTTCACTACAGCGG 5641

5634 AATACAGTACCTAGCAGGCTTATCCATCTCTGCTGAAAAACCCGCGATAGATCAATTCAT 5713
 5642 GATACAGTACCTAGCAGGCTTATCCATCTCTGCTGAAAAACCCGCGATAGATCAATTCAT 5701
 5714 GGCATTTACAGCTTCTATCATAGCCCGCTCACACCCCAAAACACCTCTCTGTTTAAAT 5773
 5702 GGCATTTACAGCTTCTATCATAGCCCGCTCACACCCCAAAATACCTCTCTGTTTAAAT 5761
 5774 CTTGGGGGATGAGGTGCTGCCAATCTGCTTCTCCACGCTGCTGAGCTTTCGTTGGG 5833
 5762 CTTGGGGGATGAGGTGCTGCCAATCTGCTTCTCCACGCTGCTGAGCTTTCGTTGGG 5821
 5834 CCGCCGCTATCGCCGAGCGGCTGTGGCAGCATAGGCTTGGGAAAGTCTCTGTTGGAAT 5893
 5822 CCGCCGCTATCGCCGAGCGGCTGTGGCAGCATAGGCTTGGGAAAGTCTCTGTTGGAAT 5881
 5894 CTTGGGGGCTATGAGGAGGCTAGCCGCGCATCTGTTGGCTTTAAGTCTATGAGCTAGG 5953
 5882 TCTGGCGGCTATGAGGAGGCTAGCCGCGCATCTGTTGGCTTTAAGTCTATGAGCTAGG 5941
 5954 CGAGTGCCTCCACCGAGGACCTGCTCAATCTTACTTCTCTGCTCACTCTCTCTGCTG 6013
 5942 CGAGTGCCTCCACCGAGGACCTGCTCAATCTTACTTCTCTGCTCACTCTCTCTGCTG 6001
 6014 CTTGCTCTGCGGCTGCTGCTGCGCAGCAATCTGCTGCGCAGCTGCTGCTGCTGCTG 6073
 6002 CTTGCTCTGCGGCTGCTGCTGCGCAGCAATCTGCTGCGCAGCTGCTGCTGCTGCTG 6061
 6074 GCTCTGCTGAGTGAATGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6133
 6062 GCTCTGCTGAGTGAATGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6121
 6134 TACGCACTATGCTGCTGAGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6193
 6122 CAGGCACTATGCTGCTGAGGAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6181
 6194 TACCACTCACTCACTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6253
 6182 TACCACTCACTCACTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6241
 6254 ATGCTCCGGCTGCTGCTGAGGATGCTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6313
 6242 TTGTTCCGGCTGCTGCTGAGGATGCTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6301
 6314 CAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6373
 6302 CAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6361
 6374 CCAACGCGGCTGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6433
 6362 CCAACGCGGCTGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6421
 6434 CCGGACACAGTCCCGGACATGCTCAAAACGCTTCCATGAGGATGCTGAGGCTGCTGAGG 6493
 6422 TGGAGACAGTCCCGGACATGCTCAAAACGCTTCCATGAGGATGCTGAGGCTGCTGAGG 6481
 6494 CTTGAGCAACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6553
 6482 CTTGAGCAACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6541
 6554 ACCCTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6613
 6542 GCTCTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6601
 6614 GAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6673
 6602 GAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6661
 6674 GTGCCCATGCCAGGTTCCGGCCCCGAAATTCCTGAGGAGTGGATGAGTGGCTGCTGCTG 6733
 6662 ATGCCCATGCCAGGTTCCAGCCCCCTGAAATTTTTCAGGAGGTTGGATGAGTGGCTGCTG 6721
 6734 CAGGTACGCTCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6793


```

QY 8954 CTCACATGAGCCACCTTGAACCTACCTCAGATCAATTGAACGACTCCATGCTCTTAGCGCAATT 3013
DB 8942 CTCACATGAGCCACCTTGAACCTACCTCAGATCAATTGAACGACTCCATGCTCTTAGCGCAATT 3001
QY 9014 TACACTCCACAGTACTCTCCAGGTAGATCAATAGGGTGGCTTCTATGCTCAGGAAACT 3073
DB 9002 TACACTCCACAGTACTCTCCAGGTAGATCAATAGGGTGGCTTCTATGCTCAGGAAACT 3061
QY 9074 TGGGGTACCAACCTTCGGAACCTGGAGACATCGGGCCAGAAGTGTCCGGCTAAGCTACT 9133
DB 9062 TGGGGTACCGCTTTCGGAGTCTGGAGACATCGGGCCAGAAGTGTCCGGCTAAGCTACT 9121
QY 9134 GTCCAGGGGGGAGGGCCGCACTTGTGCGAGATACCTCTTTAACTGGGCGAGTAAGGAC 9193
DB 9122 GTCCAGGGGGGAGGGCTGCCACTTTCGCGCAAGTACTCTTCAACTGGGCGAGTAAGGAC 9181
QY 9194 CAGAGCTTAACTCACTCCCAATCCCGCCGCGTCCAGCTGGACTTGTCTGGCTGGTGGT 9253
DB 9182 CAGAGCTTAACTCACTCCCAATCCCGCTGCCAGCTAGACTTGTCTGGCTGGTGGT 9241
QY 9254 CGTGGTTACAGCGGGGAGACATATATCACAGCTGTCTCGTCCCGACCCCGCTGGTT 9313
DB 9242 TGTGGTTACAGCGGGGAGACATATCACAGCTGTCTCGTCCCGACCCCGCTGGTT 9301
QY 9314 TCGTTGTGCTACTCCTTCTCTAGGGTGGAGATTTACTGCTCCCCAACCGATG 9373
DB 9302 CATGTTGTGCTACTCCTTCTCTAGGGTGGAGATTTACTGCTCCCCAACCGATG 9361
QY 9374 AACGGGGAGCTAACCACTCCAGGCCCTTAAGCCATTTCCTGTTTTTTTTTTT 9424
DB 9362 AACGGGGAGCTAACCACTCCAGGCCAATAGGCCAATTCCTTTTTTTTTTTT 9412

RESULT 2
US-10-259-275-39
; Sequence 39, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
; APPLICANT: Yi, Minkyung
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
; FILE REFERENCE: 265,0007 01.20
; CURRENT APPLICATION NUMBER: US/10/259,275
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US 60/171,909
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 09/747,419
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/325,236
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/338,123
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 39
; LENGTH: 9275
; TYPE: DNA
; ORGANISM: ARTIFICIAL
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding hepatitis C virus polyprotein
; OTHER INFORMATION: derived from HCV-N
US-10-259-275-39

Query Match 82.2%; Score 7885.8; DB 14; Length 9275;
Best Local Similarity 91.0%; Pred. No. 0;
Matches 8437; Conservative 0; Mismatches 817; Indels 21; Gaps 4;

QY 342 ATGAGCAGGATCCTTAACTTAAAGAAACCAACCGTAACCAACCGCCGCCACAG 401
DB 1 ATGAGCAGGATCCTTAACTTAAAGAAACCAACCGTAACCAACCGCCGCCACAG 60
QY 402 GAGCTCAAGTTCGGGGCGGTGCTCAGATCGTTGGAGTTTACCTGTTGCCGCGCAGG 461

```

```

DB 61 GACGTCAAAGTTCGCGGGCGGTGGTCAGATCGTTGGTGGAGTTTAACTGTTGCGCGCAGG 120
QY 462 GGCCTCCAGGTTGGGTGTGGCGCGACTAGGAGAGGTTTCGGAGCGGTCCCAACTCGTGA 521
DB 121 GGCCTCCAGGTTGGGTGTGGCGCGACTAGGAGAGGTTTCGGAGCGGTCCCAACTCGTGA 180
QY 522 AGGCGACAACTATATCCAAAGGCTCGCGGACCCGAGGGCAGGGCTGGGGCTCAGCCGGG 581
DB 181 AGGCGACAGGCTATATCCCAAGGCTCGCGGACCCGAGGGCAGGGCTGGGGCTCAGCCGGG 240
QY 582 TACCTTGGCCCCCTTATGGCAATGAGGGCTTGGGTGGGAGGATGCTCTCTCACC 641
DB 241 TATCTTGGCCCCCTTATGGCAATGAGGGCTTGGGTGGGAGGATGCTCTCTCACC 300
QY 642 CGCGCTCCCGGCTTGTGGGGCCCGACGACCCCGGGCTAGGTGTCGCTAACTTGGGT 701
DB 301 CGCGCTCCCGGCTTGTGGGGCCCGACGACCCCGGGCTAGGTGTCGCTAACTTGGGT 360
QY 702 AAGGTCAATCGATACCTTATCATGCGGCTTCGCGATCTCATGGGATCATTCGCTGTC 761
DB 361 AAGGTCAATCGATACCTTATCATGCGGCTTCGCGATCTCATGGGATCATTCGCTGTC 420
QY 762 GCGCGCCCTTAGGGGGCGCTGCCAGGCTTGGCAACGCTGTCGGGTTCGGAGGAC 821
DB 421 GCGCGCCCTTAGGGGGCGCTGCCAGGCTTGGCAACGCTGTCGGGTTCGGAGGAC 480
QY 822 GCGGTGAATATGCAACAGGGAACCTGCGCGGTGCTCTTCTCTATCTTCTCTTGGCT 881
DB 481 GCGGTGAATATGCAACAGGGAACCTGCGCGGTGCTCTTCTCTATCTTCTCTTGGCT 540
QY 882 CTGCTCTCTGTTGACCATCCAGCTTCCGCTTATGAAAGTCGCAACGCTGTCGGGATA 941
DB 541 CTGCTCTCTGTTGACCATCCAGCTTCCGCTTATGAAAGTCGCAACGCTGTCGGGATA 600
QY 942 TACCATGTACGAACTGCTCTCACTCAAGCATGTGTATGAGGACAGCGGACGTGATC 1001
DB 601 TACCATGTACGAACTGCTCTCACTCAAGCATGTGTATGAGGACAGCGGACGTGATC 660
QY 1002 ATGCATATCTCCCGGTGCTGCTGCTTCAAGAGGCTTAAACAGCTTCCCTGCTGGGATA 1061
DB 661 ATGCATATCTCCCGGTGCTGCTGCTTCAAGAGGCTTAACTCTCCGCTGCTGGGATA 720
QY 1062 GCGTCACTCCCAAGCTCGCGGACAGGATGACAGCTCCCACTACGACAAATACAGAC 1121
DB 721 GCGTCACTCCCAAGCTCGCGGACAGGATGACATCCCACTACGACAAATACAGAC 780
QY 1122 CAGCTCGACTGCTGCTGTTGGGAGGCTGCTTCTGCTCGCTATGATGAGTGGGGATCTC 1181
DB 781 CAGCTCGACTGCTGCTGTTGGGAGGCTGCTTCTGCTCGCTATGATGAGTGGGGATCTC 840
QY 1182 TCGGATCTATTTTCTCTCTCCAGCTGTTCACTTCTCGCTTCCGCGGACATGAGACA 1241
DB 841 TCGGATCTATTTTCTCTCTCTCTCAGCTGTTCACTTCTCGCTTCCGCGGACATGAGACA 900
QY 1242 GTGAGGACTGCAACTGCTCAATCTATCTCCGCGCAATGATAGGTCACCGATGCTGG 1301
DB 901 TCGGAGGACTGCAATGTTTCACTTACCCCGGACAGGCTCAGGTCACCGATGCTGG 960
QY 1302 GATATGATGAGAACTGCTCACTTACAGGCTTACAGGCTTACAGGCTTACAGGCTTAC 1361
DB 961 GATATGATGAGAACTGCTCACTTACAGGCTTACAGGCTTACAGGCTTACAGGCTTAC 1020
QY 1362 CCACAACTGCTGTTGGGAGCATGTTGGGCGGGGCGCCACTCGGGAGTCTTGGCGGGCTTCC 1421
DB 1021 CCACAACTGCTGTTGGGAGCATGTTGGGCGGGGCGCCACTCGGGAGTCTTGGCGGGCTTCC 1080
QY 1422 TACTATTCATGTTAGGAACTGAGGCTTAAAGTCTGATTTGTTGGGCTTACTCTTTCGCGG 1481
DB 1081 TACTATTCATGTTAGGAACTGAGGCTTAAAGTCTGATTTGATTTGATGCTACTTCTTTCGCGG 1140
QY 1482 GTTGGGGGAGACCAACAGAGGGGGGGTGGCGGCGCCACACACCTTCGCGGTTTACG 1541

```

Db 1141 GTTGAAGGACACCCCTCAACAGGGGGGAGCGCTGCCGCCCTCACCAAGGGTTGCG 1200
QY 1542 TCCCTTTTCTCATCTGGGGGCTCTCAGAAATCCAGCTTGTGAATACCAACGGCAGCTGG 1601
Db 1201 GGCTCTTTACACCTGGCGCTCTCAGAGAAATCCAGCTTATAACACCAATGSCAGTTGG 1260
QY 1602 CAAATCAACAGGAGTGCCTTAATNTGAAATGAACTCCCTCCAAATCGGTTCTTTGGCGG 1661
Db 1261 CAAATCAACAGGAGTGCCTTAATNTGAAATGAACTCCCTCCAAATCGGTTCTTTGGCGG 1320
QY 1662 CTGTTTTACGCACACAAAGTTCAACTCTCCGGGTGCCGGAGCGCAGCCAGCTGCCG 1721
Db 1321 CTGTTTACGCACACAAAGTTCAACTCTCCGGGTGCCGGAGCGCAGCCAGCTGCCG 1380
QY 1722 CCAATTGACTGGTTTGGCCCGAGGGTGGGGCCCATCACTATACCTAAGCCT---AACAGC 1778
Db 1381 TCCATTGACAAGTTTCGACACAGGATGGGTCTTATCACTTATGCTGAGCCTACAAAGAC 1440
QY 1779 TCGGATCAGAGGCTTTATGCTGGCAATTACGGCTCGACCGGTGTGTGTCGTACCGCG 1838
Db 1441 CCGGACACAGAGGCTTTATGCTGGCACTAGCCCACTCAAAATGTGGTATCGTACCTCG 1500
QY 1839 TCCAGAGTGTGCTCCAGGTATTTTCCACCCCAAGCCCTGTGTGGTGGGACCAAC 1898
Db 1501 TCGAGGTGTGCTCCAGGTATTTGCTTCAACCCCAAGTCCGTGTGCTGTGGGACAAAC 1560
QY 1899 GATCGTTCCGCTGTCCCTACGTATAGCTGGGGGAGAAATGAGACAGACGTGATGCTCCTC 1958
Db 1561 GATCGTCTCGCAACCTACGTACAGCTGGGGGAGAACGATACTGACGTGCTGCTCCT 1620
QY 1959 AACACAGCGGTTCGGCCACAGGCACTGGTTGGCTGTATCATGGATGAATAGTACGG 2018
Db 1621 AACACAGCGGTTCGGCCAGGCACTGGTTGGCTGTATCATGGATGAATAGTACGG 1680
QY 2019 TTCACTAAGACGTGCGGAGTCCCGCTGTAAACATCGGGGGTCTGGTACCCACCTTG 2078
Db 1681 TTACCAAGACGTGCGGGGCCCCCGGTGTAAACATCGGGGGTCTGGTACCAATACACCTTG 1740
QY 2079 ATCTGCCCCACGACTGCTTCCGGAACACCCGAGGCTACTTACACAAATGTGGCTG 2138
Db 1741 ACCTGCCCCACGACTGCTTCCGGAACACCCGAGGCTACTTACAAATGTGGCTG 1800
QY 2139 GGGCCCTGGTTGACACTAGTGCTAGTATGATCCCAACAGGCTTTGGCTACTCC 2198
Db 1801 GGGCCCTGGTTGACACTAGTGCTAGTATGATCCCAACAGGCTTTGGCTACTCC 1860
QY 2199 TGCACTCTCAATTTTCCATCTTAAAGTTAGGATGTATGCGGGCGGTGGAGCACAG 2258
Db 1861 TGCACTGTCAACTTTCCTTTTAAAGTTAGGATGTATGCGGGCGGTGGAGCACAG 1920
QY 2259 CTCATGCGCCATGCAATTGGACTCGAGGAGAGCGCTGTAACTTGGAGGACAGGATAGG 2318
Db 1921 CTTAAATGCTGCATGCAACTGCAACCGAGGAGAGGCTGTGAACCTTGGACGACAGGACAGA 1980
QY 2319 TCAAGACTACCGCGCTGCTGTCTTCAACAGAGTGGCAGATATGCGCCCTGTGCTTTC 2378
Db 1981 TCGAGCTACGCGCGTGTGCTCTTCAACAGAGTGGCAGGTTCTGCGCTGTGCTTTC 2040
QY 2379 ACCACCTTACCGGCTTTATCCATCTGTTTATCATCTCCATCAGAACATCTGTGGACGTG 2438
Db 2041 ACCACCTTACCGGCTTCTCTCATCTGGCTTATCATCTCCATCAGAACATCTGTGGACGTG 2100
QY 2439 CAATACCTGTACGCTGTAGGCTCAGCGTTTGTCTCTTTTCAATCAATGGAGTATATC 2498
Db 2101 CAATACCTGTACGCTGTAGGCTCAGCGTTTGTCTCTTTTCAATCAATGGAGTATATC 2160
QY 2499 CTGTTGCTTTTCTTCTCTTCCCTGGCAGCGGGGTGTGCTGCTTGTGATGATGCTG 2558
Db 2161 GTGTTGCTTTTCTTCTCTTCCCTGGGACGCGGGTCTGTGCTGCTTGTGATGATGCTG 2220
QY 2559 CTGATAGCCAGGCTGAGCGGCTTAGAGAACTTGGTGTGCTCTCATGCGGGTCCGCTG 2618
Db 2221 CTGATAGCCAGGCGGCGCTTAGAGAACTTGGTGTGCTCTCATGCGGGTCCGCTG 3360

QY 2619 GCCGAGCGCATGGTATCTCTCTTTCTTGTGTTCTTCTGCGCCGCTGGTACATTAG 2678
Db 2281 GCCGAGCGCACGGCATCTCTCTTCTCTGTGTTCTTCTGTGCGCTTGGTACATCAAG 2340
QY 2679 GGCAGCTGCTCTCTGGGGGGGTATGCTTTTATGGCGTATGGCGCTCTCTGCTC 2738
Db 2341 GGCAGCTGCTCTCTGGGGGGCATATGCTTTCTATGGCGCATGGCGCTCTCTGCTC 2400
QY 2739 CTACTCGGCTTACCAACAGAGCTTACGCTTGGACCGGAGATGCTGCATGTCGGG 2798
Db 2401 CTCTTGAATATACCAACAGAGCTTACGCTTGGACCGGAGATGCTGCATGTCGGG 2460
QY 2799 GGTGCGCTTCTTGTAGTCTGTGTTTCACTTATCCAGAGCGGAGCGACATGCAAGTG 2858
Db 2461 GGCAGCTTCTTGTGGTCTGGCATTATGACCTTGTGCCCATATTAAGGTGTCTCTC 2520
QY 2859 ACTAGGCTCATATGCTGTGTTTCAATCTTATCCAGAGCGGAGCGACATGCAAGTG 2918
Db 2521 GCTAGGCTCTATGCTGTGTTTCAATATCTTATCCAGAGCTGAGCGCACTTGCATGTG 2580
QY 2919 TGGGTCCCCCCTCAACGTTGGGGAGCGCGATGCCATCATCTCTCACTGCTGCG 2978
Db 2581 TGGGTCCCCCCTCAACGTTGGGGAGCGCGATGCCATCATCTCTCACTGCTGCA 2640
QY 2979 GTTCATCCAGAGTAAATTTTGCATCAACAACTCTCTGCTGCCCATCTCGGCCGCTC 3038
Db 2641 GTCCACCCAGAGCTAATCTTTGATATCAACAACTCTGATTCCTCACTCGGACGCTC 2700
QY 3039 ATGGTCTCAGGCTGGCATTAACAGAGTGGCTACTTCTGTGCGCGCTCAAGGCTCAT 3098
Db 2701 ATGGTCTCAGGCTGGCATTAACAGAGTGGCTACTTCTGTGCGCGCTCAAGGCTCAT 2760
QY 3099 CGTGCATGATGTTAGTCCGAAAGTCCGCGGGTCAATATGTCCAAATGCTTCTCATG 3158
Db 2761 CGTGCATGATGTTAGTCCGAAAGTCCGCGGGTCAATATGTCCAAATGCTTCTCATG 2820
QY 3159 AAGCTGGGCGCTGACAGTACGTTTATTAACCACTTTACCCCACTGCGGCACTG 3218
Db 2821 AGACTGGGCGCTGACAGGCGACGTAAGTCTATATCACTTCACTCCGCTGCGGATGG 2880
QY 3219 GCCACGCGGGCTACAGACCTTGGGTTGGGTTAGACCGCTGCTCTTCTCGGCAAG 3278
Db 2881 GCCACGCGGGCTACAGGACCTTGGGTTAGGTTAGGAGCTGCTCTTCTCTGACATG 2940
QY 3279 GAGACCAAGTTCATCCCTGGGAGCAGACCGCTGCTGTGGGACATCACTTGGGT 3338
Db 2941 GAGACCAAGTTCATCCCTGGGAGCAGACCGCGGCTGTGGGACATCACTTGGGT 3000
QY 3339 CTACCGCTCTCGCCCGAAGGGGAGGAGATATTTTGGGACCGGCTGATGCTCGAA 3398
Db 3001 CTACCTGTCTCGCCCGAAGGGAGAGATATCTCTGGGCGCGCGATAGTCTAGTA 3060
QY 3399 GGGCAAGGTGGGACTCTTGGGCGCATACGCGCTACTCCGAAACAAACGCGGGCGTA 3458
Db 3061 GGGCAAGGTGGGACTCTTGGGCGCATACGCGCTACTCCGAAACAAACGCGGGCGTA 3120
QY 3459 CTGTGTTGATCATCACTAGCTTACAGGCGGGACAGAACAGGTCCGAAAGGGAGGT 3518
Db 3121 CTGTGTTGATCATCACTAGCTTACAGGCGGGACAGAACAGGTCCGAAAGGGAGGT 3180
QY 3519 CAAGTGTCTTACCGCAACAATCTTCTGGGACCTGCACTCAACGCGCTGTGCTGG 3578
Db 3181 CAAGTGTCTTACCGCAACAATCTTCTGGGACCTGCACTCAACGCGCTGTGCTGG 3240
QY 3579 ACTGTCTACATGGCGCTGGCTCGAAGACCTTAGCGGTCGAAAGSTCCATTCACCA 3638
Db 3241 ACTGTCTACATGGCTGCTGCTCAAGACTTAGCGGCGCCAAAGGCGCAATCGCCAG 3300
QY 3639 ATGTACCAATGTAGACCTGGACCTCGTGGCTGTCAGGCGCGCCCGCGGGCGGCTCC 3698
Db 3301 ATGTACCAATGTAGACCTGGCTGCTGGCTGGCGGCGCGCGCGGGCGGCTCC 3360

QY 3699 ATGACACCATGAGCTGFGGAGCTGGGACCTTTACTTGGTCAAGAGACATGCTGATGC 3758
 Db 3361 CTGACACCATGACCTGTGGAGCTCGGACCTTTACTTGGTTACGAGACATGACAGATGTT 3420
 QY 3759 ATTCCGGTGGCCGCGGAGGAGCAGACGAGGAGAGTCTACTCTCCCGAGCCCGTCTCC 3818
 Db 3421 ATTCCGGTGGCCGCGGAGGAGCAGATAGAGGAGCTTGTCTCCCGAGCCCGTCTCC 3480
 QY 3819 TACCTGAAGAGCTCCTCGGTGGTTCATGCTTTGCCCTTCGGGCGACGCTGTGGGCTC 3878
 Db 3481 TACTTGAAGAGCTCTTCGGGTGGCCCACTGCTCTGCCCTTCGGGCGACGCTGTGGGCTC 3540
 QY 3879 TTCGGGCTGCTGTGTGACCCCGGGGTCGAGAGGGGTGGAGCTTACATACCCGTTGAG 3938
 Db 3541 TTCGGGCTGCTGTGTGACCCCGGGGTCGAGAGGGGTGGAGCTTGTCTCCCGGTTGAG 3600
 QY 3939 TCTATGAAAATACCATGCGGTCTCCGGTCTTTCAGACAACTCAACCCCGCCGCTGTA 3998
 Db 3601 TCCATGGAAAATCTATGCGGTCCCGGTCTTTCAGACAACTCAATCTCCCGCCGCTGTA 3660
 QY 3999 CCGCAGACATTCGAAGTGCGCATCTGCAAGCTCTTACTGCGAGCGGCAAGAGCACTAGA 4058
 Db 3661 CCGCAAAACATTCGAAGTGCGCATCTGCAAGCTCTTACTGCGAGCGGCAAGAGCACTAGA 3720
 QY 4059 GTGCGGCTGCTGTGTGACCCCGGGGTCCGAGAGGGGTGGAGCTTACATACCCGTTGCC 4118
 Db 3721 GTGCGGCTGCTGTGTGACCCCGGGGTCCGAGAGGGGTGGAGCTTGTCTCCCGGTTGCC 3780
 QY 4119 GCCACCTTAGGGTTGGGGCGTATATGTCGAAGGCAACGCTATCGACCTTAACATCAGA 4178
 Db 3781 GCTACCTTAGGGTTGGGGCGTATATGTCGAAGGCAACATGCTACCGACCTTAACATCAGA 3840
 QY 4179 ACTGGGTAGGACATTAACACGGGCGGCTCCATTCGACTCCACTATGGCAAGTTTC 4238
 Db 3841 ACTGGGTAGGACATTAACACGGGCGGCTCCATTCGACTCCACTATGGCAAGTTTC 3900
 QY 4239 CTTGCCACAGGTGCTGTCTGCGGCGGCTATGACATCATTAATATGCTGAGTGCAC 4298
 Db 3901 CTTGCCACAGGTGCTGTCTGCGGCGGCTATGACATCATTAATATGCTGAGTGCAC 3860
 QY 4299 TCAACTGACTCGACTACCATCTTGGGATCGGCAAGTCTCGGACCAAGCGGAGCGCT 4358
 Db 3961 TCAACTGACTCGACTACCATCTTGGGATCGGCAAGTCTCGGACCAAGCGGAGCGCT 4020
 QY 4359 GGAGCGGCTCGTCTGCTGCGCACCGCTACACTCGGGATCGGTACCGTCCACAC 4418
 Db 4021 GGAGCGGCTCGTCTGCTGCGCACCGCTACACTCGGGATCGGTACCGTCCACAC 4080
 QY 4419 CCAATATCGAGGAAATAGGCTCTCCAACTGGAGAGATCCCTTCTATGGCAAGGC 4478
 Db 4081 CCAATATCGAGGAGTGGCCCTGTGCAACTGGAGAGATCCCTTCTACGGCAAGGC 4140
 QY 4479 ATCCCAATGAGGCAATCAAGGGGGAGGCACTCAATTTCTGCCATTCGAAGAAA 4538
 Db 4141 ATCCCAATGAGGCAATCAAGGGGGAGGCACTCAATTTCTGTCACTCCAAGAGAG 4200
 QY 4539 TGTGACGAGCTCGCGCAAGCTGACAGGCTCGGACTCGGATGACGCTTACGATATACCGG 4598
 Db 4201 TGGACGAGCTTGGCGCAAGCTGTGAGGCTCGGATTCGAATCAATGCTGAGGCTATACCGG 4260
 QY 4599 GGCCTTGATGTGCTCGCTCATACCGCTTATCGGAGAGCTGTGCTGGGCAACAGAGCT 4658
 Db 4261 GGTCTTGATGTGCTCGCTCATACCGCAACGAGAGAGCTGTGCTGGGCAACAGAGCT 4320
 QY 4659 CTATATGAGGGTTTCAACCGGATTTTGAATCACTGATGAGTGCATATACATGTGCAC 4718
 Db 4321 CTATATGAGGGTTTCAACCGGATTTTGAATCACTGATGAGTGCATATACATGTGCAC 4380
 QY 4719 CAGACATGACATTCAGCTTGGATCCCACTTCCCAATGAGAGGAGCAAGTGCACCA 4778
 Db 4381 CAGACATGACATTCAGCTTGGATCCCACTTCCCAATGAGAGGAGCAAGTGCACCA 4440
 QY 4779 GACGCGTGTGCGCTCGCAACCGGAGTAGAATCGGAGGCTAGGAGTGGCATCTAC 4838

Db 4441 GACGAGTGTGCGCTCGCAGCGCGGGTAGGACTGGCAGGGGCGAGGGGGGCATATAC 4500
 QY 4839 AGGTTTGTGATCCAGGAGAAACGGCCCTCGGCAATGTTGATTTCTTGGTCTCTGTGAG 4898
 Db 4501 AGGTTTGTAACTCCGGGGGAAACGGCCCTCGGCAATGTTGATTTCTTGGTCTCTGTGAG 4560
 QY 4899 TGTATGACCGGGCTGTGCTTGTATGAGCTCAACCGCCGCTGAGACCTTCGGTTAGTTG 4958
 Db 4561 TGTATGACCGGGCTGTGCTTGTATGAGCTCAACCGCCGCTGAGACCTTCGGTTAGTTG 4620
 QY 4959 CGGCTTACCTAAATACACAGGCTTCCCGTCTGCCAGAACCATCTGGAGTTCTGGAG 5018
 Db 4621 CGGCTTACCTAAATACACAGGATTTCCCGTCTGCCAGAACCATCTGGAGTTCTGGAG 4680
 QY 5019 AGGCTTTCACAGGCTCAACCAATAGATGCCATCTTCTGTCGCCAGACTAAACAGGCA 5078
 Db 4681 AGGCTTTCACAGGCTCAACCAATAGATGCCATCTTCTGTCGCCAGACTAAACAGGCA 4740
 QY 5079 GGAGACAACCTTCTTACCTTGGTGGCATATCAAGCTACAGTGTGCGCCAGGGCTCAAGCT 5138
 Db 4741 GGAGATAACTTCTCCCTACCTGTGGCATACCAAGCCACAGTGTGCGCCAGGGCTCAGGCC 4800
 QY 5139 CCACTTCCATCTGCGGACCAATGTGGAAGTGTCTCATCGGCTGAAACCTACACTGCAC 5198
 Db 4801 CCACTTCCATCTGCGGATCAAAATGTGGAAGTGTCTCATCGGCTGAAACCTACACTGCAC 4860
 QY 5199 GGCCCAACACCCCTGCTGTATAGGCTAGGAGCGCTCCAAAATGAGGTCTATCTCACAC 5258
 Db 4861 GGCCCAACACCCCTGCTGTATAGGCTAGGAGCGCTCCAAAATGAGGTCTATCTCACAC 4920
 QY 5259 CCCATAAATAATACATCATGTCATGTCGGCTGACCTGGAGTCTGCTCACTAGCACC 5318
 Db 4921 CCCATAAATAATACATCATGTCATGTCGGCTGACCTGGAGTCTGCTCACTAGCACC 4980
 QY 5319 TGGTCTGTTAGGCGGAGTCTTGGAGCTTTGGCCGCTATCTGCTCTGACAGCAGGAGT 5378
 Db 4981 TGGTCTGTTAGGCGGAGTCTTGGAGCTTTGGCCGCTATCTGCTCTGACAGCAGGAGT 5040
 QY 5379 GTGCTCATTTGGGCGAGGATCATCTTGTCCGGAAGCGAGCTCTGCTCCGACAGGAA 5438
 Db 5041 GTGCTCATTTGGGCGAGGATCATCTTGTCCGGAAGCGAGCTCTGCTCCGACAGGAA 5100
 QY 5439 GTCTCTACAGGAGTTCGATGAGATGGAAGAGTGTGCTCTCAACTCTCTTACATCGAG 5498
 Db 5101 GTCTCTACAGGAGTTCGATGAAATGGAAGAGTGTGCTCTCAACTCTCTTACATCGAA 5160
 QY 5499 CAGGGAATCAGCTCGCGGAGGAAATCAAGCAAAAGCGCTCGGTTGTGCAAAAGGCC 5558
 Db 5161 CAGGGAATCAGCTCGCGGAGGAAATCAAGCAAAAGCGCTCGGTTGTGCAAAAGGCC 5220
 QY 5559 ACCAAGCAAGCGAGGCTCTGCTCCCGTGGAGTCCAGTGGGAGCGCTTGAGACC 5618
 Db 5221 ACCAAGCAAGCGAGGCTCGCGCTCCCGTGGAGTCCAGTGGGAGCTTTGGAGACC 5280
 QY 5619 TTCTGGGAGAGCAATGTGGAATTTTCACGCGGAATACAGTACTAGCAGGCTTATCC 5678
 Db 5281 TTCTGGGAGAGCAATGTGGAATTTTCACGCGGAATACAGTACTAGCAGGCTTATCC 5340
 QY 5679 ACTCTGCTGGAACCCCGGATGACATCATTTGATGGCAATTTACAGCTTCTATCATAGC 5738
 Db 5341 ACCCTGCTGGAACCCCGGATGACATCATTTGATGGCAATTTACAGCTTCTATCATAGC 5400
 QY 5739 CCGCTCACCAACCAAAACACCTCTGTTTAAATCTTGGGGGAGTGGGTGCTGCCAA 5798
 Db 5401 CCGCTCACCAACCAAGACACCTCTGTTTAAATCTTGGGGGAGTGGGTGCTGCCAA 5460
 QY 5799 CTGCTCTCCAGCGCTCGCTGAGCTTTCTGTTGGCGCGCGGATCGCGGAGCGGCTGTT 5858
 Db 5461 CTGCTCTCCAGCGCTCGCTGAGCTTTCTGTTGGCGCGCTGATCGCTGCTCGGCTGTT 5520
 QY 5859 GGCAGCATAGGCTTGGGAGGCTCTGCTGAGCATTTGGCGGCTATGCGGCGAGGGTA 5918

Db 5521 GCGAGCATAGTCTTGGGAAGTCTAGTGACATTCCTGGCGGCTATGGGCGAGGGGTG 5580
Qy 5919 GCGGCGCAGCTCGTGGCTTTAAAGTCAATAGAGCGGAGGTGCGCTCCACGAGGACCTG 5978
Db 5581 GCTGGCGGCTCGTGGCTTCAAGGTCAATAGAGCGGAGGCGCGCTCTGCGGAGGACCTG 5640
Qy 5979 GTCAACTTACTCTCCATGCACTCTCTCTCTGGTGGTCCCTGCTGCGGAGTCTGTGCGCA 6038
Db 5641 ATCAATTTGCTCCCTGCGCATCTCTCTCTGGTGGTCCCTGCTGCGGAGTCTGTGTGCA 5700
Qy 6039 GCAATACCTCGTGGGAGTGGGCGGAGAGGGGCTGTCAGTGGATGAACCGGCTG 6098
Db 5701 GCAATACCTCGTGGGAGTGGGCGGAGAGGGGCGGCTGTCAGTGGATGAACCGGCTG 5760
Qy 6099 ATAGCGTTGCTTGGCGGGGTAAACACGCTCTCCCTACGCACTATGTCCTGAGAGGAC 6158
Db 5761 ATAGCGTTGCTTGGCGGGGTAAACACGCTCTCCCTACGCACTATGTCCTGAGAGGAC 5820
Qy 6159 GCTGAGACGCTGTCATCAGATCTCTCTGTAAGCTTACATCACTCACTCACTGAAAGG 6218
Db 5821 GCGCAGCGGCTGTCATCAGTCTCTCTGAGGCTTCCAGGCTTACATCACTCACTGAAAGG 5880
Qy 6219 CTCACAGTGGATTAATGAGGACTGCTCTAGGCACTGCTCCGCTCGTGGCTAAGGAT 6278
Db 5881 CTCACAGTGGATTAATGAGGACTGCTCTAGGCACTGCTCCGCTCGTGGCTAAGGAT 5940
Qy 6279 GTTTGGATTTGATATGACAGCTGTTGACTGACTTCAAGACTTCAAGACTTCAAGCTC 6338
Db 5941 GTTTGGACTGGGTGTCACAGTGTGAGTGAATTCAGACTTCAAGACTTCAAGCTC 6000
Qy 6339 CTGCGCGGTTACGGGAGTCCCTTCTGTCATGCCAACCGGGGTACAGGGAGTCTGG 6398
Db 6001 CTGCGCGGTTACGGGAGTCCCTTCTGTCATGCCAACCGGGGTACAGGGAGTCTGG 6060
Qy 6399 CGGGGAGCGCATCATGCAAAACACCTGCCATGCGGAGCACAGATCGCGGACATGTC 6458
Db 6061 CGGGGAGCGCATCATGCAAAACACCTGCCATGCGGAGCACAGATCGCGGACATGTC 6120
Qy 6459 AAAAAAGGTTCCATGAGGATCGTAGGGCTAGAACCTGCAACAACAGTGGCAGCGAACG 6518
Db 6121 AAAAAAGGTTCCATGAGGATCATCGGGCGGAAACCTGCAACAACAGTGGCAGCGAAC 6180
Qy 6519 TTCCCGATCAAGATACACACGCGGACCTTGACACCTCCCGGCGGCGCAATATTC 6578
Db 6181 TTCCCGATCAAGATACACACGCGGACCTTGACACCTCCCGGCGGCGCAATATTC 6240
Qy 6579 AGGCGGCTATGGCGGCTGCTGCTGAGGAGTACGTGAGGTTACGCGTGGGGGATTC 6638
Db 6241 AAGCGGCTATGGCGGCTGCTGCTGAGGAGTACGTGAGGTTACGCGGTTGGGGATTC 6300
Qy 6639 CACTACGTGACGGGATGACCACTGACAACTGAAAGTGCCCATGCCAGTTCCGGGCCCC 6698
Db 6301 CACTACGTGACGGGATGACCACTGACAACTGAAAGTGCCCATGCCAGTTCCAGTCTCT 6360
Qy 6699 GAAATCTTACGGAGTGGATGAGTGGCTTGACAGTACGCTCGGCGGTGCAAACT 6758
Db 6361 GAGTTTTCAGGAGTGGATGAGTGGCTTGACAGTACGCTCGGCGGTGCAAACT 6420
Qy 6759 CTTCTACGGGAGGAGTCACTGCTGAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG 6818
Db 6421 CTCCTACGGGATGAGTGGATTCAGGTCGCGGCTCAATCAATACCTGTTGGGTTCAG 6480
Qy 6819 CTCCTACGGGATGAGTGGATTCAGGTCGCGGCTCAATCAATACCTGTTGGGTTCAG 6878
Db 6481 CTCCTACGGGATGAGTGGATTCAGGTCGCGGCTCAATCAATACCTGTTGGGTTCAG 6540
Qy 6879 CACATTAAGCAGAGAGCGGCTAAGGCTAGGCTGCTGAGAGGCTTCCCGCTCTTTAGCC 6938
Db 6541 CACATTAAGCAGAGAGCGGCTAAGGCTAGGCTGCTGAGAGGCTTCCCGCTCTTTAGCC 6600
Qy 6939 AGCTCATGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6993
Db 6601 AGCTCATGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6660

Qy 6994 -----ACCATGACTCCCGGAGCGGTGACCTCATCGAGGCCAACCTCTTTGTGGCGCAG 7046
Db 6661 TCCTAATACTTGTGACTCTCCGAGCTGCACTCATTTGGCGCCAACTCTCTGTGGCGCAG 6720
Qy 7047 GAGATGGCGGAAACATCACTCGCTGGAGTCAAGATTAAGGTAGTAATTTCTGGACTCT 7106
Db 6721 GAGATGGCGGAAACATCACTCGCTGGAGTCAAGATTAAGGTAGTAATTTCTGGACTCT 6780
Qy 7107 TTCGACCGCTTACGCGGAGGGGATGAGAGGAGATATCCCTCGCGGCGAGATCCCTG 7166
Db 6781 TTCGACCGCTTACGCGGAGGGGATGAGATTAAGATTAATTCATTTGCGGCGGAGATCCCTG 6840
Qy 7167 CGAAATTCAGAGGATTTCCCTCAGGCTTCCCATATGGCACGCCCGGACTCAATCCCT 7226
Db 6841 CGGAGTCCAGAGATTTCCCTCAGGCTTCCCATATGGCACGCCCGGACTCAATCCCT 6900
Qy 7227 CCACTGCTAGTCTCGGAGGACCGGACTAGTCTCTCCCTGGTGGTACACGAGTCCCA 7286
Db 6901 CCATTTGAGTCTTGGAAAGACCGGACTAGTCTCTCCCTGGTGGTACACGAGTCCCA 6960
Qy 7287 TTGCCACTACAGGCTCTCCCAATACCACTCCACGGAGAAAGAGAGCGTTGCTCTG 7346
Db 6961 TTGCCACTGTCAGGCGCTCTCCCAATACCACTCCACGGAGAAAGAGAGCGTTGCTCTG 7020
Qy 7347 ACAGAAATCCAAATGCTCTTCTGCTTGGCGGAGTCCCACTAAGACTTTTCGCTAGTCC 7406
Db 7021 ACGGACTCCACCGTCTCTTCTTGGCGGAGTCCCTACCAAACTTTCCGCGAGTCC 7080
Qy 7407 GATCTGCTGGCGGTTGATAGCGGACCGGACCGCTTCTGACTGCTGCTGCTGCTGCTG 7466
Db 7081 GAATTTGCGGCGCGGACAGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7140
Qy 7467 GGTGACAAAGATCCGAGCTTGAAGTCTGCTCTCTCATGCGGCGGCGGCGGCGGCGG 7526
Db 7141 GCGCGCAAGACTCCGAGCTGAGTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7200
Qy 7527 GGGGACCGGCTCTCAGCGGCGGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 7586
Db 7201 GGGGACCGGCTCTCAGCGGCGGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 7260
Qy 7587 GTGCTGCTGCTCAATGCTCTATAGTGACAGGCGGCTGATCAGCCATGCTGCTGCTG 7646
Db 7261 GTGCTGCTGCTCAATGCTCTATAGTGACAGGCGGCTGATCAGCCATGCTGCTGCTG 7320
Qy 7647 GAGAAATTAAGCTGCGCAATCAACCGCTTCAAGAACTCTCTTCTGCTGCTGCTGCTG 7706
Db 7321 GAAGAAAGCAAGTGCCTCAACGCTTGAAGAACTCTCTTCTGCTGCTGCTGCTGCTG 7380
Qy 7707 GTCTACGCCCAACATCCCGCAGCGCAAGCTCCGCGAGAGAGGTCACCTTTGACAGA 7766
Db 7381 GTCTACGCCCAACATCCCGCAGCGGCGGCTCCGCGAGAGAGGTCACCTTTGACAGA 7440
Qy 7767 TTGCAAGTCTGCTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7826
Db 7441 CTGAGTCTGCTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7500
Qy 7827 ACAGTTAAGCTAAGCTTCTATCTATAGAGAGGCGCTGCAAGCTGACCGGCGGCGG 7886
Db 7501 ACAGTTAAGCTAAGCTTCTATCTATAGAGAGGCGCTGCAAGCTGACCGGCGGCGG 7560
Qy 7887 GCGAAATCCAAATTTGGCTATGGGCGCAAGGAGCTCCGGAACCTATCCAGCAGGCGGCT 7946
Db 7561 GCGAAATCCAAATTTGGCTATGGGCGCAAGGAGCTCCGGAACCTATCCAGCAGGCGGCT 7620
Qy 7947 AACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8006
Db 7621 AACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7680
Qy 8007 ACCATCATGCAAAAGTGAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8066
Db 7681 ACCGTCATGCAAAAGTGAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7740

Qy	8067	GCTCGCCTTATCGTATTCCAGACCTGGAGTTCTGTATGCGAGAAGATGGCCCTTTAC	8126
Db	7741	GCCCGCTTATCGTATTCCAGATTTCGGAGTTCTGTATGCGAGAAGATGCTCTCTAC	7800
Qy	8127	GAGTGTCTCCACCCTTCCTCAGCCGTGATGGCTCCCTATACGATTTCAATACTCC	8186
Db	7801	GATGTGTCTCCACCCTTCCTCAGCCGTGATGGCTCCCTATACGGAATCCAGTACTCT	7860
Qy	8187	CCCAAGCAGCGGTCGAGTTCCTGGTGAATACCTGGAAATCAAGAATAATCCCTATGGC	8246
Db	7861	CCCGGCAGCGGTCGAGTTCCTGGTAAAGCCTGGAAATCAAGAATAATCCCTATGGC	7920
Qy	8247	TTCTCATATGACACCCGCTTTTTCACACGCTCACTGAGTGTGACATTCGTGTGAG	8306
Db	7921	TTCTCATATGACACCCGCTTTTTCACACGCTCACTGAGATGACATTCGGTGTGAG	7980
Qy	8307	GAGTCAATTACCAATGTGTGACTTGGCCCCGAGGCCAGACAGGCCATAAGTTCGCTC	8366
Db	7981	GAGTCAATTACCAATGTGTGACTTGGCCCCGAGGCCAGACAGGCTATAAATCGCTC	8040
Qy	8367	ACAGAGCGCTTTACATCGGGGTCCTCGACTAACTCAAAAGGCGAGACTCGGTTAT	8426
Db	8041	ACAGAGCGCTTTATACGGGGTCCCTCGACTAACTCAAAAGGCGAGACTCGGTTAT	8100
Qy	8427	CGCCGCTCCCGCGCAAGTGGCGTCTGACGACTAGCTCGGTAATACCTCACAATTTAC	8486
Db	8101	CGCCGCTCCCGCGAGCGCGTCTGACGACTAGCTCGGTAATACCTCACAATTTAC	8160
Qy	8487	TTGAGAGCCATGACGCTCTGAGCTGCAAAAGCTCCAGAGTGCACAGTCTCGTGAAC	8546
Db	8161	TTGAAAGCCTTCGCGCTCTGAGCTGCAAAAGCTCCAGAGTGCACAGTCTCGTGAAC	8220
Qy	8547	GGAGACACCTTGTGCTTATCTGTGAAAGCGCGGAACCCAGGAGGATCGCGGCGCCTA	8606
Db	8221	GGGAACACCTTGTGCTTATCTGCAAGCGCGGAACCCAGGAGGATCGCGGAGCTTA	8280
Qy	8607	CGAGCCTTCAGGAGCGCTATGACTAGTATTCGCGCCCCCGGGGATTCGCGCCCAACCA	8666
Db	8281	CGAGCTTCAGGAGCGCTATGACTAGTATTCGCGCCCCCGGGGATTCGCGCCCAACCA	8340
Qy	8667	GAATACACACTGGAGCTGATAACATCATGTTCTTCCAAATGTGTCAGTCGGCAGCATGCA	8726
Db	8341	GAATACACTTGGAGTTGATAACATCATGTTCTTCCAAATGTGTCGGCAGCATGCA	8400
Qy	8727	TCCTGGCAAAAGGTATCTACTTACCCGTGACCCACACCCGCCCTTGCAACGGGCTGCG	8786
Db	8401	TCCTGGCAAAAGGTGTACTACCTCACTCGCGATCCACACCCGCCCTTGCAACGGGCTGCG	8460
Qy	8787	TGGGAGACACTAGACACTCCAACTCAACTCTTGGCTAGGCAATATCATCTGATGCG	8846
Db	8461	TGGGAACACTAGACACTCCAGTTAATCTTGGCTAGGCAATATCATCTGATGCG	8520
Qy	8847	CCACCCATGGGCAAGGATGATCTGATGACTCACTTTTCTCCATCTTCTAGTCAA	8906
Db	8521	CCACCCATTATGGGCAAGGATGATCTGATGACCATTTCTTCTCCATCTTCTAGTCAA	8580
Qy	8907	GAGCACTTGAAAGCCCTGGATGTCAATCTACGGGCTTGCTACTCCATTGAGCCA	8966
Db	8581	GAGCACTTGAAAGCCCTGGATGCAAACTACGGGCTTGCTACTCCATTGAGCCA	8640
Qy	8967	CTTGACCTACTCAGATCATTTGAACCACTCCATGGTCTTAGCGATTTCACATCCACAGT	9026
Db	8641	CTTGACCTACTCAGATCATTTGAAGGACTCCATGGTCTTAGCGATTTCACATCCATAGT	8700
Qy	9027	TACTTCCAGTGAGATCAATAGGGTGGCTTCATGCTCAGGAACTTGGGTACCAACC	9086
Db	8701	TACTTCCAGTGAGATCAATAGGGTGGCTTCATGCTCAGGAACTTGGGTACCAACC	8760
Qy	9087	TTCCGAACCTTGGAGACATCCGGCCAGAAAGTCCGGCTTAAGTACTCTTCCAGGGGG	9146
Db	8761	TTCCGAGTCTGGAGACATCCGGCCAGGAGCTCCGGCTTAACTACTCTTCCAGGGGGG	8820
Qy	9147	AGGCGGCCACTTCTGGCAGATACCTCTTTAACTGGGCAATGAGGCAACAGCTTAACTC	9206

Db	8821	AGGCCGCCACTTGC	GGCAATAACCTCTTCAACTGGCAGTAAGACCAAGCTCAAACTC	8880
Qy	9207	ACTCCAATCCGGCCGGTCC	CAGCTGGACTGTCTGCGCTGGTTTCGTGCGTGGTTACAGC	9266
Db	8881	ACTCCANTCCGGCTGGTCC	CAGTTAATCCGGCTGGTTCTGTTGCTGGCTACAGC	8940
Qy	9267	GGGGGACATATATAC	AGCCCTGTCTCGGCCGACCCCGCTTCGTTCTGTGCTTA	9326
Db	8941	GGGGGACATATATAC	AGCCCTGTCTCGTGGCCGACCCCGCTGGTTCAATGCTGTGCCTA	9000
Qy	9327	CTCCTACTTTCTGTAGGGT	TAGGCATTTACTGCTCCCGAACCGATGAACGGGGAGCTAA	9386
Db	9001	CTCCTACTTTCTGTAGGGT	TAGGCATCTACTTGTCCCGCAATCGATGAACGGGGAGCTAA	9060
Qy	9387	CCACTCCAGGCC--TTAAG	GCATTTCTGTT	9445
Db	9061	ACACTCAGGCCCAATAG	GCATTTCTGTT	9120
Qy	9446	TTTTTTTTCTTCCTTCCT	CTTTTTTTTCCCTTCTTTTTTCCCTTCT-----TTAATGGT	9500
Db	9121	TTTTTTTTTTTTTTTTTTTT	TTTTTTTCCCTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATGT	9180
Qy	9501	GGTCCATCTTAGCCCT	PAGTCAGGGCTAGCTGTGAAGGTCCGTGAGCGCGCATGACTGCA	9560
Db	9181	GGCTCCGCTTACGCCCT	TAGTCAAGGCTAGCTGTGAAGGTCCGTGAGCGCGCATGACTGCA	9240
Qy	9561	GAGAGTCTGATTA	CTGGCCCTCTCTGCAGATCACTGT	9595
Db	9241	GAGAGTCTGATTA	CTGGCCCTCTCTGCAGATCACTGT	9275

```

RESULT 3
US-10-189-359-13
; Sequence 13, Application US/10189359
; Publication No. US20040039187A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, Annette
; APPLICANT: SANGAR, DAVID V.
; APPLICANT: LEMON, STANLEY M.
; TITLE OF INVENTION: Chimeric GB Virus B (GBV-B)
; FILE REFERENCE: UTSG:258US
; CURRENT APPLICATION NUMBER: US/10/189,359
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: 10/189,359
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentia Ver. 2.1
; SEQ ID NO 13
; LENGTH: 9599
; TYPE: DNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342) ..(9377)
US-10-189-359-13

```

20

QY 181 GACGACGGGCTCTTCTTGGATCAACCGCTCAATGCTGGAGATTGGGGCTGCCCC 240
Db 181 GACGACGGGCTCTTCTTGGATCAACCGCTCAATGCTGGAGATTGGGGCTGCCCC 240
QY 241 GCGAGACTGCTAGCCGAGTAGTGTTCGGTCGCGAAAGGCTTGTGTACTGCTGATAGG 300
Db 241 GCGAGACTGCTAGCCGAGTAGTGTTCGGTCGCGAAAGGCTTGTGTACTGCTGATAGG 300
QY 301 GTGCTTGCAGTAGTCCCGGGAGGTCTCTGTAGACCGTGCACCATGAGCAAGATCTTAAC 360
Db 301 GTGCTTGCAGTAGTCCCGGGAGGTCTCTGTAGACCGTGCACCATGAGCAAGATCTTAAC 360
QY 361 CTCRAAGAAAACCAACCTTAACCAACCGCGGCCCAAGAGCTCAAGTCCCGGGCG 420
Db 361 CTCRAAGAAAACCAACCTTAACCAACCGCGGCCCAAGAGCTCAAGTCCCGGGCG 420
QY 421 GTGGTCAGATGCTGTGTGGAGTTAAGTCTGTCGCGGAGGGGCCCGAGTGGGTGTC 480
Db 421 GCGGTCAAGATGCTGTGTGGAGTTAAGTCTGTCGCGGAGGGGCCCGAGTGGGTGTC 480
QY 481 GCGGACTAGGAAGGTTCCGAGCGGTGCAACCTGCTGGAGGCGACCACTATCCCAA 540
Db 481 GCGGACTAGGAAGGTTCCGAGCGGTGCAACCTGCTGGAGGCGACCACTATCCCAA 540
QY 541 AGGCTCGCCGACCCGAGGSCAGGCTGGCTCAGCCGCGGTACCGTGGCCCTCTATG 600
Db 541 AGGCTCGCCGACCCGAGGSCAGGCTGGCTCAGCCGCGGTACCGTGGCCCTCTATG 600
QY 601 GCAATGAGGGCTGCGGGTGGGAGGATGCTCTGTCAACCGCGGCTCCCGGCTAGT 660
Db 601 GCAATGAGGGCTGCGGGTGGGAGGATGCTCTGTCAACCGCGGCTCCCGGCTAGT 660
QY 661 GGGGCCCCACGACCCCGGGTAGTTCGGTAACTTGGTAAGTCAATGATACCCCTTA 720
Db 661 GGGGCCCCACGACCCCGGGTAGTTCGGTAACTTGGTAAGTCAATGATACCCCTTA 720
QY 721 CATGGGGCTCGCCGATCTCATGGGTACATTCCTCTGTCGGGCGCCCGCTAGGGGG 780
Db 721 CATGGGGCTCGCCGATCTCATGGGTACATTCCTCTGTCGGGCGCCCGCTAGGGGG 780
QY 781 CTGCCAGGGCTTGGCAACGGTGTCCGGTCTTGGAGGAGGGCGTGAATGCAACAG 840
Db 781 CTGCCAGGGCTTGGCAACGGTGTCCGGTCTTGGAGGAGGGCGTGAATGCAACAG 840
QY 841 GGAATCTCCCGTGTGCTCTTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db 841 GGAATCTCCCGTGTGCTCTTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
QY 901 TCCGAGCTTCCGCTTATGAAGTGGCAACGCTGTCGGGATATACCATGTCAGBAAGCT 960
Db 901 TCCGAGCTTCCGCTTATGAAGTGGCAACGCTGTCGGGATATACCATGTCAGBAAGCT 960
QY 961 GCTCCTACTAAGATTTGTATGAGGAGCGGAGTGAATCATGATCTCCCGGGTGG 1020
Db 961 GCTCCTACTAAGATTTGTATGAGGAGCGGAGTGAATCATGATCTCCCGGGTGG 1020
QY 1021 TGCCTGTGTTTCAGAGGTAACAGCTCCGCTGCTGGTAGCGCTACTCCAGGCTG 1080
Db 1021 TGCCTGTGTTTCAGAGGTAACAGCTCCGCTGCTGGTAGCGCTACTCCAGGCTG 1080
QY 1081 CGGCCAGGAATGCGAGGCTCCCACTACGACAAATACGAGCGCAAGTGGAGTGTGCT 1140
Db 1081 CGGCCAGGAATGCGAGGCTCCCACTACGACAAATACGAGCGCAAGTGGAGTGTGCT 1140
QY 1141 GGAAGGCTGTTTCTGCTCGCTATGATGCGGGGATCTCTGGGATATATTTTCCTG 1200
Db 1141 GGAAGGCTGTTTCTGCTCGCTATGATGCGGGGATCTCTGGGATATATTTTCCTG 1200
QY 1201 TCTCCAGCTGTTTCACTTCTCGCTGCGGCGATGAGAGTGCAGGATGCAACTGCT 1260
Db 1201 TCTCCAGCTGTTTCACTTCTCGCTGCGGCGATGAGAGTGCAGGATGCAACTGCT 1260
QY 1261 CAATCTATCCCGGCGATGATCAGGTCAACCGCATGGCTTGGGATATGATGATGACTGGT 1320

Db 1261 CTATCTATCCCGGCGATATTAACGGGTCAATGCGATGGCATGGATGATGATGACTGGT 1320
QY 1321 GACCTACAAAGCCCTAGTGTGTCGAGTGTCTCCGATGCCAAGAGTGTCTGAGACA 1380
Db 1321 GACCTACAAAGCCCTAGTGTGTCGAGTGTCTCCGATGCCAAGAGTGTCTGAGACA 1380
QY 1381 TGGTGGCGGGGCGCCACTGGGAGTCTTGGGGGCTTGGCTACTATTCTATCTGAGGGA 1440
Db 1381 TGGTGGCGGGGCGCCACTGGGAGTCTTGGGGGCTTGGCTACTATTCTATCTGAGGGA 1440
QY 1441 ACTGGGCTAAGGCTTCTGATTTGGGCTTACTCTTGTCCGGGCTTGAAGGGGAGACCCACA 1500
Db 1441 ACTGGGCTAAGGCTTCTGATTTGGGCTTACTCTTGTCCGGGCTTGAAGGGGAGACCCACA 1500
QY 1501 CGAGGGGAGGCTGCGCGGCCACACCTCCGGGCTTCAAGTCCCTTTCTCATCTGGGG 1560
Db 1501 CGAGGGGAGGCTGCGCGGCCACACCTCCGGGCTTCAAGTCCCTTTCTCATCTGGGG 1560
QY 1561 CGTCTCAGAAAATCCAGCTTGTGAATACCAACGGCAGCTGGCACATCAACAGGACTGCC 1620
Db 1561 CGTCTCAGAAAATCCAGCTTGTGAATACCAACGGCAGCTGGCACATCAACAGGACTGCC 1620
QY 1621 TAAATTCAGTGAATCTCCCTCAAACTGGGTCTTGTCCGGCTTGTATGCGACACAAGT 1680
Db 1621 TAAATTCAGTGAATCTCCCTCAAACTGGGTCTTGTCCGGCTTGTATGCGACACAAGT 1680
QY 1681 TCAACTGCTCCGGGTGCGCGGCGCATGCGAGCTGCGCGGCCCATGACTGGTTCGGCC 1740
Db 1681 TCAACTGCTCCGGGTGCGCGGCGCATGCGAGCTGCGCGGCCCATGACTGGTTCGGCC 1740
QY 1741 AGGGTGGGGCCCCATCACCTATATAAGCTTAACAGCTCGCATCAGAGGCTTATGCT 1800
Db 1741 AGGGTGGGGCCCCATCACCTATATAAGCTTAACAGCTCGCATCAGAGGCTTATGCT 1800
QY 1801 GGCATTTAGCGCTCGACCGGTGTGGTTCGATACCGGCTGCGAGTGTGTGCTCAGTGT 1860
Db 1801 GGCATTTAGCGCTCGACCGGTGTGGTTCGATACCGGCTGCGAGTGTGTGCTCAGTGT 1860
QY 1861 ATTGTTTCAACCCCAAGCCCTGTTGGTGGGACCAAGATCTGTCGGTGTCCCTAGT 1920
Db 1861 ATTGTTTCAACCCCAAGCCCTGTTGGTGGGACCAAGATCTGTCGGTGTCCCTAGT 1920
QY 1921 ATAGTGGGGGGAATGAGACAGACCTGATGCTCTCAACAAACGCGTCCGCGACAAG 1980
Db 1921 ATAGTGGGGGGAATGAGACAGACCTGATGCTCTCAACAAACGCGTCCGCGACAAG 1980
QY 1981 GCAACTGTTCCGGTGTACATGAGTGAATAGTACTGGTTCACCTAAGACCTGCGAGGTC 2040
Db 1981 GCAACTGTTCCGGTGTACATGAGTGAATAGTACTGGTTCACCTAAGACCTGCGAGGTC 2040
QY 2041 CCGCTGTAAACATCGGGGGGCTCGGTAAACGCGACCTGATCTGCTCCCAACGAGCTGCTTC 2100
Db 2041 CCGCTGTGTATCGGAGGGGTGGGCAACACCTTGTCTGCTCCCAACGAGCTGCTTC 2100
QY 2101 GGAAGCACCCCGAGGCTTACTTACAAAATGTGGCTCGGGGCGCTGCTGACACCTAGT 2160
Db 2101 GGAAGCACCCCGAGGCTTACTTACAAAATGTGGCTCGGGGCGCTGCTGACACCTAGT 2160
QY 2161 GCTTACTAGTACTACCATACAGGCTTGGCATCTACCGCTGCACTCTCAATTTTTCATCT 2220
Db 2161 GCTTACTAGTACTACCATACAGGCTTGGCATCTACCGCTGCACTCTCAATTTTTCATCT 2220
QY 2221 TTAAGGTTAGGATGATATGGGGGGGTGGAGACAGGCTCAATGCGGATGCAATTTGA 2280
Db 2221 TTAAGGTTAGGATGATATGGGGGGGTGGAGACAGGCTCAATGCGGATGCAATTTGA 2280
QY 2281 CTCGAGGAGGCTGCTTAATCTTGGAGACAGGATAGGTCAAGAACTCAGCCGCTGCTGC 2340
Db 2281 CTCGAGGAGGCTGCTTAATCTTGGAGACAGGATAGGTCAAGAACTCAGCCGCTGCTGC 2340
QY 2341 TGTCTTCAACAGAGTGGCGATATCTCCCTGTGCTTTCACCACTTACCGGCTTATCCA 2400

Db 2341 TGTCCACACACAGTGGCAGGTCTCTTCCTGCTCTTTTCAGCACCTCGCCAGCTTGTCCA 2400
Qy 2401 CTGGTTTTCATCTCCATCAGAAACATCGTGGAGTGCATACCTGTACGGTGTAGGCT 2460
Db 2401 CCGGCTCATCCACTCCACAGAACATTTGAGCTGCAGTACTTTGTACGGGTAGGCT 2460
Qy 2461 CAGCGTTTCTCTCTTTGCAATCAAAATGGAGTACATCTGTGTCTTCTCTCTCTCTG 2520
Db 2461 CAAAGCATCCGCTCTCTGGGCAATAAGTGGAGTACGCTCTCTCTCTCTCTCTCTCTG 2520
Qy 2521 CAGACGCGCGGTGTGCT 2580
Db 2521 CAGACGCGCGGTGTGCT 2580
Qy 2581 CTTTAGAGAACTTGTGTCT 2640
Db 2581 CTTTGGAGAACT 2640
Qy 2641 CTTTCT 2700
Db 2641 CTTTCT 2700
Qy 2701 CGTATGCTTTTATGGCGTATGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
Db 2701 TCTAGCCCTCTACGGGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
Qy 2761 CTTACGCTTTGACCGGGAGATGGCTGATCGTGGCGGGTGGCTCTCTCTCTCTCTCTCTCT 2820
Db 2761 CATACGCACTGACACCGAGGTGGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2820
Qy 2821 TATTCTTGACCTTGTACCATACTACAAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
Db 2821 TGGCGCTGACTCTCTCGCCATATTAACAAGCGTATATCAGCGTGTGTGCTGTGTGTGTG 2880
Qy 2881 AATACTTTATACACAGACCGAGGCGCACATCCAAAGTGTGGTCTCTCTCTCTCTCTCTCT 2940
Db 2881 AGTATTTCTGACAGAGTAGAGCGCACTCCAGTGTGGTCTCTCTCTCTCTCTCTCTCTCT 2940
Qy 2941 GGGAGGCGCGATGCGATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3000
Db 2941 GGGGCGCGCGATGCGATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3000
Qy 3001 ACATCAACAACT 3060
Db 3001 ACATCAACAACT 3060
Qy 3061 CGAGAGTGCCTGCT 3120
Db 3061 TTAAAGTCCCTACT 3120
Qy 3121 AAGTCCCGGGGTCAATATGTCAAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3180
Db 3121 AGATAGCGGAGGTCAATACGTCGCAATGCGCATCATCAAGTTAGGGCGCTTACTGGCA 3180
Qy 3181 CGTACGTTTATAACCATCTTACCCACTGCGGACTTGGGCCCAACGCGGCTTACGAGACC 3240
Db 3181 CCTATGTATTAACCATCTTACCCCTCTCTGAGACTTGGCGCAACAGGCTTGGAGATC 3240
Qy 3241 TTGCGGTGGCGGTAGAGCCGCT 3300
Db 3241 TGGCGGTGGCTGTGAAACGAGTCT 3300
Qy 3301 GAGCAGACCGCTGGGTGGGACATCATCTTGGGTCTACCCGTCTCCGCGCGAAGG 3360
Db 3301 GGGCAGATACCCCGCGTGGGTGACATCATCAACGGGTCTGCGCTCTCTCTCTCTCTCTCTCT 3360
Qy 3361 GGAAGAGATATTTTGGACCGGCTGATAGTCTGGAAGGCAAGGTTGGGACTCTCTCTCTCT 3420
Db 3361 GCCAGAGATCTGCTTGGGCGAGCGGACGGATGCTCTCAAGGGTGGAGTTGCTCTG 3420
Qy 3421 CGCCCATCAAGCGCTTCTCCAAACAAACGCGGGGCTACTTGGTTGCATCATCAAGCC 3480
Db 3421 CGCCCATCAAGCGCTTCTCCAAACAAACGCGGGGCTACTTGGTTGCATCATCAAGCC 3480

Qy 3481 TCACAGCCCGGCAAGAAACAGGTCCGAAGGGAGGTTCAAGTGGTTTCTTACCGCAACAC 3540
Db 3481 TGAATGCGCGGCAAGAAACAGGTCCGAAGGGTGAAGTCCAGATCGTGTCACTGCTACCC 3540
Qy 3541 AATCTTCTCTGGCGAATGCAATCAACCGCGTGTCTGAGCTGTCTACATGGCGCTGCT 3600
Db 3541 AAACTCTCTGGCAACGTGATCAATCGGGTATCTGGAATGTCTAGCAACGGGCGGAA 3600
Qy 3601 CGAAGACCTAGCGGTCCAAAGGTCCAAATCAACCAAAATGTACCAACCAATGTAGACCTG 3660
Db 3601 CGAGGACATCGCATCAACCAAGGTCTCTGTATTCAGATGTATACCAATGTAGACCAAG 3660
Qy 3661 ACCTCGTCTGGCTGGCAGCGCCCGCGCGCGCTCCATGACACCACTGACAGCTGTGCA 3720
Db 3661 ACCTGTGGCTGGCGCGCTCTCTCAAGTTCCTCATGACACCTGTACCTGCGCT 3720
Qy 3721 GCTCGGACCTTACTTGGTCAAGAGACATGCTGATGTCTATTCGGTGGCGCGGAGCG 3780
Db 3721 CTTGGACCTTACTTGGTCAAGAGACATGCTGATGTCTATTCGGTGGCGCGGAGCTG 3780
Qy 3781 ACAGCAGGGAGTCTACTCTCCCGCAGCGCGCTCTCTACTCAAAAGGTCTCTCGGCTG 3840
Db 3781 ATAGAGGGGTAGCTGCTTTCGCCCGCGCCATTTCTACTTGAAGGCTCTCTCGGCG 3840
Qy 3841 GTCCATTTGCTTGGCTTTCGGGCACTGCTGGCGCTCTTTCGGGCTCTCTGCTGCAACC 3900
Db 3841 GTCCGCTGTGTGCGCGCGGACACGCGTGGCGCTTATTCAGGCGCGCTGTGCAACC 3900
Qy 3901 GGGGCTCGGAAAGCGGTGACTTCAATCCGCTGTACCGCAGACATTCCAAGTGGCAC 3960
Db 3901 GTGAGTGGCTTAAAGCGGTGACTTATTCCTGTGGAGAACCTTAGGACACACCTAGAT 3960
Qy 3961 CTCGCTCTTTCACAGACAACTCAACCGCGCTGTACCGCAGACATTCCAAGTGGCAC 4020
Db 3961 CCCGCTGTTCACGAGCAACTCTCTCTCCACAGCAGTGCCTCCAGAGCTTCCAGTGGCC 4020
Qy 4021 ATCTGCACTCTCTTCTGGCAGCGCAAGACCAAAAGTGCCTGCTGCTGATGCAAGCC 4080
Db 4021 ACCTGCACTCTCTCCACCGGCAAGAGTAAAGACCAAGTCCCGCTGCTGATGCAAGCC 4080
Qy 4081 AAGGTCAGAGTCTCTCTGCTGAACTCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4140
Db 4081 AGGCTCAAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4140
Qy 4141 ATATGTCAGAGCACAGGTATCGACCTCAACCTAACTCAGAACTGGGGTAAGACCACTTACCA 4200
Db 4141 ACATGTCAGAGCACAGGTATCGACCTCAACCTAACTCAGAACTGGGGTAAGACCACTTACCA 4200
Qy 4201 CCGGCGCTCCATTAAGTATCTCACTCTATGCAAGTCTCTTGGCAGCGTGGCTGCTGCTG 4260
Db 4201 CTGCGAGCCCATCACTGCTCTCACTCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4260
Qy 4261 GGGCGCTCTGACATCATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Db 4261 GAGTGTCTTATGACATATATATTTGTGACGAGTCCCACTCCACGATGCACTCACTCACT 4320
Qy 4321 TGGCATCGGACAGTCTCTGACCAAGCAGAGACTGCGGGGCGAGACTGTTGTGCTG 4380
Db 4321 TGGCATCGGACAGTCTCTGACCAAGCAGAGACTGCGGGGCGAGACTGTTGTGCTG 4380
Qy 4381 CCACGCTACACTCTCGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4440
Db 4381 CCACTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4440
Qy 4441 TGTCCCACTGAGAGACTCCCTTCTTATGCAAGCCATCCCTTATGAGGCCATCAAG 4500
Db 4441 TGTCCCACTGAGAGACTCCCTTCTTATGCAAGCCATCCCTTATGAGGCCATCAAG 4500
Qy 4501 GGGGAGGCT 4560
Db 4501 GGGGAGGCT 4560

6721 Dd EGGTGCGCTACACAGGTTTGGCCGCCCTTGCAGGCCCTTGGTGGGGAGGAGGTATCAT 6780
6781 Qy TCCAGGTGGGCTCAACCAATCTGGTGGGTGCGAGCTCCATGCGAGCCCGAACCGG 6840
6781 Dd TCAGAGTAGAATCCACAGAGTACCGGGTGGGTCGAATTAATTGCGAGCCCGAACCGG 6840
6841 Qy ACCTAACAGTGTCTTCCATGTCTCACCGATCCCTCCCAATTAACAGCAGAGAGCGCTA 6900
6841 Dd ACCTAGCCGTGTGAGCTCCATGTCTCACTGATCCCTCCCAATTAACAGCAGAGGGCGCG 6900
6901 Qy AGGTAGGCTGGCTGAGAGGTCTCCCTCTCTTACCGAGCTCACTAGTACCGAGTGT 6960
6901 Dd GGAGAGGTTGGCGAGAGGTCAACCCCTTCTATGCGCAGCTCTCGGTAGCCAGCTGT 6960
6961 Qy CTGCGCTCTTCTTGAAGGCGCATGACTACCCATACCTACCTCCCGAGCGCTGACTCA 7020
6961 Dd CGGTCCATCTCTCAAGGCAACTTGCACCGCCAAACATGACTCCCTGACCGCGAGCTCA 7020
7021 Qy TCGAGCCCAACCTCTTGTGGCGGAGGATCGTGGCGGGAATCAATCACTCCCGTGGAGTCAG 7080
7021 Dd TAGAGCTAACCTCTCTGTGGAGGAGAGATGGCGGCAACATCAACAGGTTGAGTCAG 7080
7081 Qy AGNATAGGTAGTAATCTTGGACTCTTTCGAACCGCTTCAACGGGAGGGGATGAGGG 7140
7081 Dd AGAACAAAGTGGTGAATCTTGGACTCTCTCGATCCGCTTGTGGCAGAGGAGATGAGCGGG 7140
7141 Qy AGATATCCCTGCGCGGAGATCCTCGAATAATCCAGGAAGTTCCCTCAGCGTTGCCCA 7200
7141 Dd AGGTCTCCGTACTGTCAGAAATCTCGGAGTCTCGAGATTCGCGGGCCCTGCCCCG 7200
7201 Qy TATGGGCAACCCCGGACTCAATCTTCCACTGTCTAGAGTCTTGGAAAGGACCGGACTAG 7260
7201 Dd TCTGGCGGCGCGGACTCAACCCCGCTAGTAGAGACGTGAGAAAGCCCTGACTAGG 7260
7261 Qy TCCTCGGTGTGACGAGTGCCTATCCCTACCGAGCTCCAGAGCTCTCCAAATACCACTC 7320
7261 Dd AACCACTGTGTGTCATGGTGCCTCGGTACCACTTCAAGTCCCTCTCTGTCCTCCCG 7320
7321 Qy CACGGAAGAGGAGGAGTGTCTTGCAGAAATCCAAATGTGTCTTGTCTTGGCGGAGC 7380
7321 Dd CTGGAAAGAGCTACGGTGTCTTCAAGAAATCAACCCATCACTGCTTGGCGGAGC 7380
7381 Qy TCGGCAATAGACCTTCGGTAGTTCGGATTCGTCGGCGGTGTAGGGCAACCGGACCG 7440
7381 Dd TTGCCAACAAAGTTTGGAGTCTCTCAACTTCCGCAATTAACCGGCAATACGACAA 7440
7441 Qy CCCTTCTGACCTGGCTCCGACGAGTGCACAAAGATCCGAGTTTGGTGTGACTCTCT 7500
7441 Dd CATCTCTGAGCCGCCCTTCTGCTGTCGCCCGGAGTCCGAGCTTGAATCTATCTT 7500
7501 Qy CCATGCCCCCTTGAAGGGGAGCCGGGAGCCCGATCTTCAGCGAGGCTTGTGTCTA 7560
7501 Dd CCATGCCCCCTTGGAGGGGAGCTGGGATCCGGATCTCAGCGACGGGTCTATGTCGA 7560
7561 Qy CCGT---GAGTGAGGAGCTAGTGAGATCTCTGCTCTGCTCAATGTCTATACGTGGA 7617
7561 Dd CCGTCAAGTACTGGGCGGACACGGAAGATCTGCTGCTCAATGTCTATTCCTGGA 7620
7618 Qy CAGCGCCCTGATCACCCATGGCTCGGAGGAAAGTAAAGCTGCCATCAACCCGTTGA 7677
7621 Dd CAGCGCACTCTGTCACCCCTGGCTCGGAGAAAGAAAGAACTGCCATCAACGACTGA 7680
7678 Qy GCAACTCTTCTGCTGCTCAACAAATGTGTCTACGCCCAACATCCCGAGCGCAAGCC 7737
7681 Dd GCAACTCTTCTGCTCAACAAATGTGTCTACGCCCAACATCCCGAGCTGCTGCC 7740
7738 Qy TCCGCGAGAGAGTCACTTTGACAGATTGCAAGTCCCTGGATGATCACTTCGCGGAGC 7797
7741 Dd AAAGCGAGAGAAAGTCACTTTGACAGACTGCGAGTCTGCGAGCCATTAACGAGCG 7800
7798 Qy TACTCAAGGAGATCAAGCGGAGCGCTCCAGTCTTAAGGCTTAAGCTTCTATCTATAGG 7857
7801 Dd TGCTCAGAGGTTCAAGACGCGCGTCAAAAGTGAAGGCTACTTCTATCCGTAGAGG 7860

7858 Qy AGGCTGCAAGCTGACGCCCCCACAATTCGGCCAAATCCAAATTTGGCTATGGGCAAGG 7917
7861 Dd AAGCTTGAGCTTGACGCCCCCACAATTCAGCCAAATCCAAATTTGGCTATGGGCAAGG 7920
7918 Qy ACCTCGGAAACCTATCCAGAGGCGGTTAACCACTCCGCTCCGTGCGGAGGACTGCG 7977
7921 Dd ACCTCGGTTGCACTCCAGAAAGCCGTTAGCCCACTCACTCCGTTGGAAAGCCCTTC 7980
7978 Qy TGGAAAGCACTGAAACACCAATTCACACCACTATGCGAAAGTGAAGGTTTCTGCG 8037
7981 Dd TGGAAAGCACTGTAACCAATTAACCACTATGCGAAAGTGAAGGTTTCTGCG 8040
8038 Qy TCCAAACAGAAAGGAGGCGCAAGCAGCTTCGCTTATCGTATTCCTCAACCTGGAG 8097
8041 Dd TCCAGCTGAGAAAGGGGTCGTAAGCAGCTCTCTCATCTGTTCCCTCCAGCTGGGCG 8100
8098 Qy TCGTGATTCGAGAGAGTGGCCCTTTACGAGGTGGTCTCCACCTTCTCAGCGCGTGA 8157
8101 Dd TCGCGTGTGGAGAGATGGCCCTGTACGAGCTGGTTAGCAAGCTCCCTGCGCGTGA 8160
8158 Qy TGGGCTCTCTCATACGATTTCAATCTCCCAAGCAGCGGTTCGAGTTCCTGTTGAATA 8217
8161 Dd TGGGAGCTCTCTACGATTCCTCACTCAACAGGACAGCGGTTGAATTCCTGTTCAAG 8220
8218 Qy CCGGAAATCAAGAAATGCGCTTATGGCTTCTCATATGACACCGCTGTTTGAATCAA 8277
8221 Dd CCGTGAAGTCCAAAGAGACCCCGATGGGGTTCTCGTATGATACCGCTGTTTGAATCCA 8280
8278 Qy CGGTCACTGAGAGTCACTTCGTTGAGGAGTCAATTTACCAATGTTGACTTGGGCC 8337
8281 Dd CAGTCACTGAGAGCACTTCGTTGAGGAGGCAATTTACCAATGTTGACTTGGGCC 8340
8338 Qy CCGAGGCGAGACAGCGCCATAAGGTGCTCACAGCGGCTTTACATCGGGGTTCCTTGA 8397
8341 Dd CCGAGGCGGCTGGCCATCAAGTCCCTCACTAGAGGCTTATGTTGGGGCCCTCTTA 8400
8398 Qy CTAACTCAAGAGGAGAACTCGCGTTATCGCGTGGCGCAAGTGGCTGCTGACGA 8457
8401 Dd CCAATTCAGGGGGGAAACTCGCGCTACCGAGGTGCGCGAGCGCGCTACTGACAA 8460
8458 Qy CTAGCTGGGTAAATCCCTCACTGTTACTTGAAGGCCACTCGAGCTGTTCGAGTTCGAA 8517
8461 Dd CTAGCTGTGTAACACCTCTCACTTGTCTATCAACAGGCCCGGAGCGCTTCGAGCGGAG 8520
8518 Qy AGCTCCAGGACTGCAAGATGCTGTGAAAGGAGAGCACTTGTCTTATCTGTGAAAGG 8577
8521 Dd GCTCCAGGACTGCAAGATGCTGTGAAAGGAGAGCACTTGTCTTATCTGTGAAAGG 8580
8578 Qy CCGGAAACCCAGAGGATGGGGCGGCTTACGAGCTTCAACGAGGCTTATGACTAGTAT 8637
8581 Dd CCGGGGTCCAGGAGGACCGCGGAGCTTACGAGGCTTCAACGAGGCTTATGACAGGACT 8640
8638 Qy CCGCCCCCCCCGGGATCCGCCCAACAGAAATAGACCTGAGGCTGATAACATCATGTT 8697
8641 Dd CCGCCCCCCCCGGGAGCCCAACAGGAGTATGAGGCTTATGAGGCTTATGAGGCTTAT 8700
8698 Qy CTTCAATGTGAGTTCGCGAGAGTCTGCGAAAGGAGTATCTACCTCAACCGG 8757
8701 Dd CTTCCAAACGCTCAGTTCGCGAGAGTCTGCGAAAGGAGTATCTACCTTACCGG 8760
8758 Qy ACCCAACCGGCTTTCAGCGGCTGCTGGAGAGCACTAGACACACTCCAAATCAACT 8817
8761 Dd ACCCTACACCGGCTTTCAGCGGCTGCTGGAGAGCACTAGACACACTCCAAATCAACT 8820
8818 Qy CTTGGCTAGGCAATATCATGATGCGCCCAACCGCTTGGGCAAGGATGTTCTGATGA 8877
8821 Dd CTTGGCTAGGCAATATCATGATGCGCCCAACCGCTTGGGCAAGGATGTTCTGATGA 8880
8878 Qy CTCACCTTTTCTCCTTCTAGCTCAAGAGCACTTGAAGGCTTGGATGTTGATGA 8937
8881 Dd CCACTTTTCTAGCTTCTCAAGAGGAGTTCAGCTTGAAGGAGTTCAGCTTGAAGGAGT 8940


```

QY 8938 TCTACGGGCTTGTCTACCTCCATGAGCCACTTGAGCTTACCTCAGATCAATTCAGACATCC 8997
Db 8941 TCTACGGGCTTGTCTACCTCCATGAGCCACTTGAGCTTACCTCAGATCAATTCAGACATCC 9000
QY 8998 ATGGTCTTAGCGCATTTATACATCCACAGTACTCTCCAGGTGAGATCAATAGGGTGGCTT 9057
Db 9001 ATGGCTCAGGCGCATTTTCACTCCACAGTACTCTCCAGGTGAGATCAATAGGGTGGCGG 9060
QY 9058 CATGCTCAGAAATCTCGGGTACACCTTGGCACTGCGAAGCTGCGGCGCAGAGTG 9117
Db 9061 CATGCTCAGAAATCTCGGGTACACCTTGGCACTGCGAAGCTGCGGCGCAGAGTG 9120
QY 9118 TCCGCGCTAAGCTACTGTCACAGGGGGAGGGCGCCACTTGTGCGAGATACCTCTTTA 9177
Db 9121 TCCGCGCTAAGCTACTGTCACAGGGGGAGGGCGCCACTTGTGCGAGATACCTCTTTA 9180
QY 9178 ACTGGGAGTAAAGCAAGCTTAACTCACTCCATCCCGGCGCGCTCCAGCTGGACT 9237
Db 9181 ACTGGGAGTAAAGCAAGCTTAACTCACTCCATCCCGGCGCGCTCCAGCTGGACT 9240
QY 9238 TGTCTGGCTGGTTCGCTGCTGTACAGGGGGAGACATATATACAGACCTGCTCGTG 9297
Db 9241 TGTCTGGCTGGTTCGCTGCTGTACAGGGGGAGACATATATACAGACCTGCTCGTG 9300
QY 9298 CCGACCCCGCTGGTTCGCTGCTGTACCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9357
Db 9301 CCGACCCCGCTGGTTCGCTGCTGTACCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9360
QY 9358 TGTCTCCCAACGATGAGGGGAGCTAACCACTCCAGGC-CTTAAGCAATTCCTGTTT 9415
Db 9361 TGTCTCCCAACGATGAGGGGAGCTAACCACTCCAGGC-CTTAAGCAATTCCTGTTT 9420
QY 9417 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9476
Db 9421 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9480
QY 9477 TTTCTTTTCTCTTTTAAAGTGGCTCCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 9536
Db 9481 TTTCTTTTCTCTTTTAAAGTGGCTCCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 9540
QY 9537 AGTTCGCTGAGCGCATCTGACAGAGAGTGTGATCTGGCTCTCTGAGATCATGT 9595
Db 9541 AGTTCGCTGAGCGCATCTGACAGAGAGTGTGATCTGGCTCTCTGAGATCATGT 9599

```

RESULT 4

```

US-09-238-076-5
; Sequence 5, Application US/09238076
; Patent No. US20020102540A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESS: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-238-076-5

Query Match 65.3%; Score 6264; DB 9; Length 12980;
Best Local Similarity 78.4%; Pred. No. 0;
Matches 7565; Conservative 0; Mismatches 2030; Indels 53; Gaps 3;

QY 1 GCGAGCCCCCTGATGGGGGCGACACTCCACATGAATACCTCCCTCTGAGGAATCTACTG 60
Db 1 GCGAGCCCCCTGATGGGGGCGACACTCCACATGAATCACTCCCTCTGAGGAATCTACTG 60
QY 61 TCCTTCAGCAGAAAGCGTCTAGCCATGCGTGTAGTATGAGTGTGTCAGCCTCCAGGAC 120
Db 61 TCCTTCAGCAGAAAGCGTCTAGCCATGCGTGTAGTATGAGTGTGTCAGCCTCCAGGAC 120
QY 121 CCCCCCTCCCGGGAGAGCCATAGTGTCTGCGGAAACCGGTGAGTACACCGGAATTCGCAG 180
Db 121 CCCCCCTCCCGGGAGAGCCATAGTGTCTGCGGAAACCGGTGAGTACACCGGAATTCGCAG 180
QY 181 GACGACCGGTCCTTCTTTGGATCAACCCGCTCAATCCCTGGAGATTGGGCGTGCCTCC 240
Db 181 GACGACCGGTCCTTCTTTGGATCAACCCGCTCAATCCCTGGAGATTGGGCGTGCCTCC 240
QY 241 GCGAGACTGTAGCCGAGTGTGTTGGTTCGCGAAAGCCCTTGTGGTACTGCCCTGATAGG 300
Db 241 GCGAGACTGTAGCCGAGTGTGTTGGTTCGCGAAAGCCCTTGTGGTACTGCCCTGATAGG 300
QY 301 GTGCTTCGAGTGTCCCGGAGGTCCTCGTAGACCGTGCACCATGAGCACGAATCTTAAC 360
Db 301 GTGCTTCGAGTGTCCCGGAGGTCCTCGTAGACCGTGCACCATGAGCACGAATCTTAAC 360
QY 361 CTCAAGAAACCAACAGTAACACACCGCCGCCACAGGAGCTCAAGTTCCTCCGCGG 420
Db 361 CTCAAGAAACCAACAGTAACACACCGCCGCCACAGGAGCTCAAGTTCCTCCGCGG 420
QY 421 GTGCTCAGATCTGTTGGTGGATTTTACCTGTTGCGCGCAGGGGCCCGAGTGGGTGTC 480
Db 421 GCGGTCAATCTGTTGGTGGATTTTACCTGTTGCGCGCAGGGGCCCGAGTGGGTGTC 480
QY 481 GCGGACTTAGGAAGCTTCCGAGCGGTGCGCAACCTCGTGAAGCGGACAACTTATCCCAA 540
Db 481 GCGGAGCAGGAAGACTTCCGAGCGGTGCGCAACCTCGAGGTAGACGTCAGCTATCCCA 540
QY 541 AGGCTCCGACCCCGAGGAGGCGCTGGGCTGAGCCCGGATACCTTGGCCCTCTATG 600
Db 541 AGGCACTCGGCGCGAGGAGGAGGACCTGGGCTCAGCCCGGCTACCTTGGCCCTCTATG 600
QY 601 GCAATGAGGCTTGGGCTGGGAGGATGGCTCTCTGTCAACCCCGGCTCCCGGCTAGTT 660
Db 601 GCAATGAGGCTTGGGCTGGGAGGATGGCTCTCTGTCTCCCGTGGCTCTCGGCTAGCT 660
QY 661 GGGGCCCCCAGGACCCCCCGGCTAGGTGCGGTAACTTTGGGTAAAGGTATCATCATCCCTTA 720
Db 661 GGGGCCCCCAGGACCCCCCGGCTAGGTGCGGTAAAGGTATCATCATCATCCCTTA 720
QY 721 CATCGGCTTCGCGGATCTCATGCGGTACATTCGCTCGTGGGCGCCCCCTTAGGGGCG 780
Db 721 CGTGGGCTTCGCGGATCTCATGCGGTACATTCGCTCGTGGGCGCCCCCTTAGGGGCG 780

```

QY	781	CTGCCAGGCGCTTGGCACACGGGTCTCCGGGTCTCTGGAGACGGCGTGAACACTATGCAACAG	840
Db	781		
QY	841	GTGCCAGGCGCTTGGCGCATGGCGTCCGGGTCTGGAAGCGCGTGAACACTATGCAACAG	900
Db	841		
QY	901	TCCACGCTTCGGCTTATGAAGTCGCAACGTGTCGGGATATACCAATGTCACGAACGACT	960
Db	901		
QY	961	GTCCCAACTCAAGCATGTGTATAGGCGACGGACGTGATCATGCAATATCCCGGTGGG	1020
Db	961		
QY	1021	TGCGCTGTGTTACGAGGGGTAAACAGTCCCGTTGCTTGGGTAGCGGCTCACTCCCAACGCTCG	1080
Db	1021		
QY	1081	CGGCCAGGAATGCACGCTCCCACTACGACAAATACGACGCCACGTCGACTTCTCTGTTG	1140
Db	1081		
QY	1141	GGACGCTGCTTCTGCTCGCTATGTACGTGGGGATCTCTCGGATCATTTTCTCTCG	1200
Db	1141		
QY	1201	TCCTCCACAGTGTTCACCTTCTCGCCTCGCCGGCATAGACAGTGCAGGACTGCAACTGCT	1260
Db	1201		
QY	1261	CAATCTATCCGGGCATGTATCAGTACACGCGATGCGTTGGGATATGATGAACCTGGT	1320
Db	1261		
QY	1321	CACCTACACAGCCCTAGTGTGTCGAGTGTCTCCGATCCCAAGCTCCCAAGCTGCTGGACA	1380
Db	1321		
QY	1381	TGCTGGCGGGGCCACTGGGGAGTCTGGCGGCGCTTGCTACTATTCATGTTAGGGA	1440
Db	1381		
QY	1441	ACTGGGCTAAGGTTCTGATTGTCGCTACTCTTTGCGGCGTTGACGGGAGACCCACA	1500
Db	1441		
QY	1501	CGACGGGAGGGTGGCCGGCCACACCACTCCGGGTTACAGTCCCTTTTCTCATCTGGGG	1560
Db	1501		
QY	1561	CGTCTCAGAAATCCAGCTTGTGAATACCAACGCGCGTGGCATCAACAGGACTCGCC	1620
Db	1561		
QY	1621	TAAATTGCAATGACTCCCTCCAACTGGGTTCTTTGGCGGCTGTTTACSCACACAAGT	1680
Db	1621		
QY	1681	TCACTGTTCGGGTGCCGAGCGCATGGCGAGCTGCCGCCCACTGACGTGTTGCGCC	1740
Db	1681		
QY	1741	AGGGTGGGGCCCATCACTTATCTAAGCCCTAACAGCTCCGATCAGAGGCTTTATTGCT	1800
Db	1741		
QY	1801	GGCAATACGCGCTCGACCGTGTGGTGTGTCGACCGCGTCCGAGGTGTGGTCCAGTGT	1860
Db	1801		

1861	QY	ATTTGTTACCCCAAGCCCTGTTGTGTGTGGGACCAACCGATCGTTCCCGGTGTCCTTACGT	1920
1861	Db		1920
1861	QY	ATTTGTTCACTCCCAAGCCCGGTGTGTGTGGGAAACGACGACAGGTGCGGGCGCGCTACCT	1920
1861	Db		1920
1921	QY	ATAGCTGGGGGAGAAATGAGACAGACGTGATGCTCCTCAACAAACAGCGGTCCGCCACAAG	1980
1921	Db		1980
1921	QY	ACAGCTGGGFTGCAATGATATAGGATGCTTCGTCCTTAAACAACACAGGCCACCGCTGG	1980
1921	Db		1980
1981	QY	GCACCTGTTGGCTGTATACATGGATCAATATGATCTGGGTTTCACTAAGACGTGCGGAGGTC	2040
1981	Db		2040
1981	QY	GCAAATGTTGCGTTGTATCTTGATGAACTCAACTGSGATTCAACAAAGTGTGCGAGGCG	2040
1981	Db		2040
2041	QY	CCCCGTGTACATCCGGGGGTGCGTAAACGACACTTGATCTGCCCCACAGCACTCTCC	2100
2041	Db		2100
2041	QY	CCCCTTGTGTATCCGAGGGGTGGCCAAACAACCTTGTCTGCCCCACTGATTTGTTCC	2100
2041	Db		2100
2101	QY	GGAAACACCCGAGGCTACTTACACAAAATGTGTGCTCGGGCCCTGGTTGACCACTAGGT	2160
2101	Db		2160
2101	QY	SCAAGCATCCGGAAGCCACATACTCTCGGTGGGCTCGGTCCCTGGATATACCCAGGT	2160
2101	Db		2160
2161	QY	GCCTAGTAGACTTACCAATACAGGCTTTGGCACTATCCCTGCACTCTCAATTTTCCAICT	2220
2161	Db		2220
2161	QY	GCATGGTGCACTACCCGATATAGCTTTGGCACTATCTCTTGTACCAATCAATACCAATAT	2220
2161	Db		2220
2221	QY	TTAAGGTTAGGATGTATGTGGGGGGCGTGAGACACAGGCTCAATCCGCATCAATTTGA	2280
2221	Db		2280
2221	QY	TCAAAGTCAAGATGTACGTGGGAGGGGTGACACAGGCTGGAAGCGGCGCTCAACTGA	2280
2221	Db		2280
2281	QY	CTGAGGAGAGCGCTGTACTTTGGAGGACAGGATAGTTCAGAACTCAGCCCGCTGCTCG	2340
2281	Db		2340
2281	QY	CGCGGGCGAAAGCTGTGATCTTGGAGACAGGACAGGTCCGAGTCAAGCCATTTGCTCG	2340
2281	Db		2340
2341	QY	TGTCTACACAGAGTGGCAGATACTGCCCTGTGCTTACACACCTTACCGGCTTTATCCA	2400
2341	Db		2400
2341	QY	TGTTCAACACAGTGGCAGGTCCCTCCGTGTTCTTTCAGACCTGCCAGCCTTGTCCTCA	2400
2341	Db		2400
2401	QY	CTGTTTGTATCCATCCCATACAAACATCGTGAAGTGCAATACCTGTAGGTGTAGGGT	2460
2401	Db		2460
2401	QY	CGGGCTCATCCACTCCACCAGAAATTTGTGACGTGCAGTACTTGTACGGGGTAGGT	2460
2401	Db		2460
2461	QY	CAGCGTTGTCTCTTTGCAATCAATGGAGGTACATCTGTGTCTTTCTCTCTCTCTCTG	2520
2461	Db		2520
2461	QY	CAAGCATCGCGTCCCTGGGCAATTAATGGAGATACGTGTTCTTCCTGTTCTCTCTCTG	2520
2461	Db		2520
2521	QY	CAGACGGCGGTGTGCTGCTTGTGAATGATGCTGTGATAGCCCAAGGCTGACGCGG	2580
2521	Db		2580
2521	QY	CAGACGGCGGTCTGCTCTGCTTGTGATGATTACTCATATCCCAAGCGGAGGCGG	2580
2521	Db		2580
2581	QY	CCTTAGAAACTTGGTGTCTCAATGGCGCTCCGTGGCCGAGGCGCATGTTTCTCT	2640
2581	Db		2640
2581	QY	CTTTGGAGAACCTCGTAATCAATTCAGATATCCCTTGGCCGGGACGACGTTCTGTGT	2640
2581	Db		2640
2641	QY	CCTTTCTGTGTTCTTCTGCGCGCTGTGTACATTAAGGCGAGGCTGGTCTCTGGGCGG	2700
2641	Db		2700
2641	QY	CCTTCTCTGTGTTCTTCTGCTTTCGTGTATCTGAAGGGTAGTGGGTGCCCGGAGCGG	2700
2641	Db		2700
2701	QY	CGTATGCTTTTATGGGGTATAGGCGCTGTCTCTGTCTCTACTCTGGGTTTACCAACGAG	2760
2701	Db		2760
2701	QY	TCTACGCTTTCTACGGGATGTGCCCTCTCCCTGTCTCTGCTGGCGTTGCCCTACGCGG	2760
2701	Db		2760
2761	QY	CTTACGCTTTGACCGGAGATCGGTGCATCTGTGCGGGGTGCGGTTCTTTGTAGTCTGG	2820
2761	Db		2820
2761	QY	CATACGACATGGACACGGAGTGGCGCGTCTGTGTGGCGGTGTGTTCTTGTGCGGGTAA	2820
2761	Db		2820
2821	QY	TATTTCTGACCTTGTACCAATATCAAAAGTTTCTCACTAGGCTCATATGTTGTTTAC	2880
2821	Db		2880
2821	QY	TGGCGCTGATCTGTGCCCATTTACAAGCGCTACATCAGCTGGTGCATGTGGTGGCTTC	2880
2821	Db		2880
2881	QY	AATACTTTTATCAACAGAGCCGAGGCCACATGAAAGTGTGGTCCCGCCCTCAACGCTC	2940
2881	Db		2940
2881	QY	AGTATTTTCTGACACAGATGAAGCGCAATGACACGTGTGGGTTCGCCCTCAACGTC	2940
2881	Db		2940
2941	QY	GGGAGGCGCGGATGCCAATCATCTCCTCTCAGCTGTGCGGTTTCATCCAGAGTTAATTTTG	3000
2941	Db		3000

QY 5161 TGTGGAAGTGTCTCATAGCGCTGAACCTACACTGACAGCGGCCAAACACCCCTGCTGTATA 5220
 Db 5161 TGTGGAAGTGTGTATCCCGCTTAACCCACCTCCATGAGGCCAAACACCCCTGCTGTATA 5220
 QY 5221 GGCTAGAGACCGTCCAAATAGGTCAATCTCAACACCCCAATACTAAATACATATGG 5280
 Db 5221 GACTGGCGCTGTTCAGAAATGAAGTCAACCTGACGACCCCAATACCAAAATACATATGA 5280
 QY 5281 CATTGATGCTGGCGCTGACCTGAGGTGTCTACTAGCACTGGGTGCTGTAGGCGGAGTCC 5340
 Db 5281 CATGCACTGCGCGGACCTGGAGTGTCTACAGAGCACTGGGTGCTGTAGGCGGCTGTC 5340
 QY 5341 TTGCAGCTTTGGCGCGCTATGCTGCTGACGACAGGCACTGTGGTCAATGTGGGAGGATCA 5400
 Db 5341 TGCGTGTCTGGCGCGCTATGCTGCTGACAGGCTGCGTGGTCAATGTAGTGGGAGGATTC 5400
 QY 5401 TCTTGTCCGGGAAGCCAGCTGTGCTTCCGACAGGGAAGTCTCTACAGAGGTGTGATG 5460
 Db 5401 TCTTGTCCGGGAAGCCGCAATTATACCTGACAGGGAAGTCTCTACAGAGGTGTGATG 5460
 QY 5461 AGATGGAGAGTGTGCTCTCAACTCTTACATCGAGAGGGAATGACAGCTGCGCGAGC 5520
 Db 5461 AGATGGAGAGTGTCTCTCAGCACTTACCGTACATCGAGAGGGAATGATGCTGCTGAGC 5520
 QY 5521 AATTCAAGCAAAAGCGCTCGGCTGTGTTGCAACCGGCCACCAAGCAAGCGAGGCTGCTG 5580
 Db 5521 AGTTCAAGCAAGAGGCTCGGCTCTGCTGACAGCCGCTGCGGCAAGCAGAGGTTATCA 5580
 QY 5581 CTCCGCTGGTGGAGTCCAAAGTGGCGAGCCTTGAGACTCTTCTGGGCGAAGCAATGGA 5640
 Db 5581 CCCCTGCTGTCAGACCAACTGGCAGAACTCGAGGCTCTTCTGGGCGAAGCAGATGGA 5640
 QY 5641 AATTCAATCAGCGGAATACAGTACAGTACAGGCTTATCCACTCTGCTGGAACCCCGCGA 5700
 Db 5641 AATTCAATCAGTGGATACAACTACTTGGGCGCTGTCAACGCTGCTGTGTAACCCCGCA 5700
 QY 5701 TAGCATCAATTGATGGCAATTTACAGCTTCTATCACTAGCCGCTCACCACCCCAAAACCC 5760
 Db 5701 TTGCTTCAATTGATGGCTTTTACAGCTGCGGTCAACAGCCCACTAACCACTGGCCAAACCC 5760
 QY 5761 TCCTGTTTAACTCTGGGGGATGGGTGGTGGTGGCGCAACTGCTCTCCACAGGCTGGT 5820
 Db 5761 TCCTCTTTCAACATATGGGGGGTGGGTGGTGGTGGCGAGCTCGCGCCCCCGGTCCGCTA 5820
 QY 5821 CAGCTTTTGTGGGCGCGCATCGCCGAGCGGCTGTGGCAGCATAGGCTTGGGAAGG 5880
 Db 5821 CCGCTTTGTGGGCGCTGGCTTACTGCGCGCGCATCGGCAGCGTTGGACTGGGGAAGG 5880
 QY 5881 TGCTCGTGGACATCTTGGCGGGCTATGGGGCAGGGGTAGCCGGCGCACTGTGGCTTTA 5940
 Db 5881 TCCTCGTGGACATCTTTCAGGGTATGGCGGGCGGTGGCGGGAGCTCTTGTAGCCTTCA 5940
 QY 5941 AGGTCAATGAGCGGAGTGGCTCCACCGAGGACCTGGTCAACTTACTCCCTGCCATCC 6000
 Db 5941 AGATCAATGAGCGGTAGTCCCTCCACGGAGGACCTGGTCAACTTCTGCTGCCCGCATCC 6000
 QY 6001 TCTCTCTGTGTGCTGTGCTGCGGGTCTGTGTCGAGCAATATCTGCTGCGGCACTGG 6060
 Db 6001 TCTCGCTTGGAGCTTGTAGTGTGGTGTCTGCGAGCAATATCTGCGCGGCACTGG 6060
 QY 6061 GCCCGGAGAGGGGCTGTGAGTGGATGAACCGGCTGATAGGTTCCCTTGGCGGGGTA 6120
 Db 6061 GCCCGGCGAGGGGCGAGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 6120
 QY 6121 ACCAGCTCTCCCTACGACATATGTGCTGAGAGGACGCTGAGCACTGTCTACTCAGA 6180
 Db 6121 ACCATGTTTCCCGACGACATATGTGCGCGAGAGGATGAGCGCGCGCTGACTGCCA 6180
 QY 6181 TCTCTCTAGCTTACCATCACTCACTGCTGAAGCGGCTCCACAGTGGATTAATGAGG 6240
 Db 6181 TACTCAGCAGCTCACTGTAACCCAGCTCTTGGGCGGACTGATCAGTGGATAGCTCGG 6240

QY 6241 ACTGCTCTACGCCATGCTCCGCGCTCGTGGCTAAGGATGTTTGGATTGGATATGACGG 6300
 Db 6241 AGTGTACCACTCCATGCTCGGTTCTTGCTAAGGACATCTGGGACTGGATATGCGAGG 6300
 QY 6301 TGTGACTGACTTCAAGACCTGGCTCCAGTCCAAACTCTCTGCGCGGTTTACCGGAGTCC 6360
 Db 6301 TGTGAGCGACTTTTAAAGCTTGGCTGAAAGCCAGCTCATGCCACAACCTGCTGGATTC 6360
 QY 6361 CTTTCTCTCATGCCAAACCGGGGTACAAGGAGTCTGCGGGGGGAAGGCAATCATGCAAA 6420
 Db 6361 CTTTCTCTCTGCTCCAGCGCGGTATAGGGGGGTCTGCGAGAGAGACGCAATATGACACA 6420
 QY 6421 CACCTGCGCATCGGAGACACAGATCGCGGACATGTCAAAAACGGTTCCTAGAGATCG 6480
 Db 6421 CTGCTGCGCACTGTGGAGCTGAGATCACTGGACATGTCAAAAACGGGAGATGAGATCG 6480
 QY 6481 TAGGGCTAGAACCTTGCAGCAACAGTGGCAACGTTCCCCATCAACGCAATACACCA 6540
 Db 6481 TCGTCTTAGGACCTTGCAGGAACATGTGGAGTGGAGCTTCCCATTAACGCTACACCA 6540
 QY 6541 CGGGACTTTGCAACCCCTCCCGCGGCCAACTATTTCAGGGCGCTATGCGCGGTGGTGG 6600
 Db 6541 CGGGCCCTGTACTCCCTTCTGCGCGCAACTATAAGTTCTGCGCTGTGGAGGGTGTCTG 6600
 QY 6601 CTGAGAGTACGTGGAGTTCAGCGTGTGGGATTTCCACTACGTGACGGGCATGACCA 6660
 Db 6601 CAGAGGAATACGTGGAGTAAAGCGGGTGGGGACTTCCACTACGTATCGGTATGACTA 6660
 QY 6661 CTGACAACTAAAGTCCCATGCCAGTTCCGCGCCCCCGAATCTTCAAGGAGGTGGATG 6720
 Db 6661 CTGACAACTTAAATGCCCTGTGCCAGATCCCATGCGCCCGAATTTTCAAGATTTGAGG 6720
 QY 6721 GAGTGGGTGACACAGTACGTCGCGCTGTGAAACCTCTTTTACGGGAGGAGTCACTG 6780
 Db 6721 GGGTGGCGCTTACATAGTTTGGCGCCCTTGTGAAGCCCTTGTGCGGGAGGAGTATCAT 6780
 QY 6781 TCCAGTGGGCTCAACCAATACTTGGTGGGTGGGAGTCCCATGCGAGCGCGGACCGG 6840
 Db 6781 TCAGATGAGGACTTCCAGAGTACCGGTTGGGGTGGGAACTTAACCTTGGAGCGCGGAG 6840
 QY 6841 ACCTAACAGTGTCTTCTCATGCTCAGCGATCCCTCCCACTTACAGAGAGACGCTCA 6900
 Db 6841 ACCTAGCGGTGTGAGTCTCATGCTCACTGATCCCTCCCACTTAAACAGAGAGCGGCGG 6900
 QY 6901 AGCGTGGCTGGTACAGGCTTCCCGCTCTTTTACCGAGCTCATGCTAGCGAGTGTGT 6960
 Db 6901 GGAAGAGTGGGAGAGGCTCAACCGCTTCTATGGCCAGCTCTCGCGGAGCGAGTGT 6960
 QY 6961 CTGCGCTTCTTGAAGGCGACATGCACTACCCACCATGACTCCCGGAGCGCTGACTCA 7020
 Db 6961 CCGTCCCATCTCTAAGGCACTTGGCACCGCCAACTATGACTCCCTGAGCGCGAGTCA 7020
 QY 7021 TCGAGGCCAACCTCTTGTGGCGGAGAGATGGCGGAAACATCATCTGCGGTGGAGTCA 7080
 Db 7021 TAGAGGCTAACCTCTGTGGAGGAGAGATGGCGGCAACATCACAGGGTGTGAGTCA 7080
 QY 7081 AGAATAGGTAGTAAATCTGGAATCTTTTCAACCGCTTACCGGAGGGGAGTGAAGGG 7140
 Db 7081 AGAACAAGTGGTGAATCTGGAATCTTTTCAACCGCTTGTGGAGAGGAGTGAAGGG 7140
 QY 7141 AGATATCCGTGCGCGGAGATCTCGAATAATCCAGGAAGTTCCTCAGCGCTTGGCCA 7200
 Db 7141 AGGTCTCCGTACCGCGAATAATCTCGGAGTCTCGAGATTCGCGCGGCTTGGCGG 7200
 QY 7201 TATGGGCAACCGGACTACAACTCTCTGCTAGAGTCTGGAAGGAGCCCGAGTACG 7260
 Db 7201 TTTGGGCGCGGAGTCAACACCCCGCTAGTAGAGCTGGAAGGAGTGAAGTACG 7260
 QY 7261 TCCCTCGGTGTACAGGATGCGCATGCGCATACCAAGGCTCTTCAATACCACTC 7320
 Db 7261 AACCACTGTGTCTCATGGTGGCGGCTACCACTCCACGCTCCCTCTGTGCTCTCGG 7320
 QY 7321 CACGGGAAGAGGAGCGGTTGCTCTGACAGATCCCAATGTGTCTTCTGCTTGGCGGAGC 7380

Db 7321 CTGGAAAAGCGTACGGTGGTCTTCCAGCAATCAACCCCTAATCTACTGCTTGGCGAGC 7380
QY 7381 TCGCCACTAAGACCTTTGGTAGTTCGGATCGTGGCGGTTGATAGCGGCAACGGGACCG 7440
Db 7381 TTGGCAACCAAGATTTTGGCAGCTCTCTCAACTTCCGGCAATTAACGGGCGACAATACGACAA 7440
QY 7441 CCCTTCTGACCTGGCTCCGACGACGGTGAATAAGGATCCGAGCTTGAGTCTGACTCTCT 7500
Db 7441 CATCTCTGAGCGCGCCCTTCTGGCTGGCCCCCGGACTCCGACGTTGAGTCTGATTTCTT 7500
QY 7501 CCATGCCCCCTTGAAGGGGAGCGCGGACCCCGATCTCAGCGAGCGGCTCTTGGTCTA 7560
Db 7501 CCATGCCCCCTTGGAGGGGAGCGCTGGGATCTCGGATCTCAGCGAGCGGCTCATGCTGA 7560
QY 7561 CCCTTCTGAGGAGGCTAGTGAAGATGCTGCTCTCAATGCTCTTAATGCTCTTCTGGA 7620
Db 7561 CGGTCACTAGTGGGCGGACACGGAAGATGCTGCTGCTCAATGCTCTTCTGGA 7620
QY 7618 CAGCGCCCTGATCACGCGATGCGCTGGGAGGAAAGTAAGCTCCCATCAACCCGTTGA 7677
Db 7621 CAGCGCACTCGTCAACCCGCTGGCTGGGAAGAAACAAAACTGCCCATCAACGCACTGA 7680
QY 7678 GCAACTCTTGTGCTGCTCACCAACATGCTACGCGCACAAATCCCGCAGCGCAAGCC 7737
Db 7681 GCAACTCTTGTACGCCATCACAATCTGGTGATTCACCACTTCACGCGAGTCTTGCC 7740
QY 7738 TCGCGCAGAGAAGGTCACTTTGACAGATGCGAAGTCTGGATGAATCAATPACGGGACG 7797
Db 7741 AAAGGCAGAGAAGATCACATTTGACAGACTGCAAGTCTTGGACAGCCATTAACAGGACG 7800
QY 7798 TACTCAGGAGATGAAGCGAGCGCTCCACAGTTAGGCTTAAGCTCTATCTATAGAG 7857
Db 7801 TGCTCAAGGAGGTCAAGACGCGCGCTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAG 7860
QY 7858 AGGCTCTCAAGCTGACGCGCCCACTTCGGCAAAATCCAAATTTGGCTATAGGGCAAGG 7917
Db 7861 AAGCTTGACGCTGAGCGCCCACTTCACGCAATCCAAAGTTGGCUAUGGGCAAAAG 7920
QY 7918 ACCTCGGAACTATPCAGCAGGCGCGTTAACCAATCCGCTCCGCTGGTGGAGACTTGC 7977
Db 7921 ACCTCGCTTGCATGCCAGAAAGCGGTAGCCACATCAACTCCGCTGGAAAGACCTTC 7980
QY 7978 TGGAGACACTGAACACCAATTCGACACCACTATCGACCAAAAGTGAAGTCTTCTGCG 8037
Db 7981 TGGAGACAGTGAACCAATTAACACTACATCATGSCCAAGAACGAGTTTCTGCG 8040
QY 8038 TCCAAACAGAGAAGGGGCGCGCAAGCCAGCTGCGCTTATCGTATTCACACCTTGGAG 8097
Db 8041 TTCAGCTGAGAAGGGGGTGGTAAGCCAGCTGCTCATGCTGTTCCCGGACCTGGGCG 8100
QY 8098 TTCGTGTATGCGAGAGATGGCCCTTACGAGGTGCTCCACCTTCTCAGGCGCTGA 8157
Db 8101 TGCCTGTGCGAGAAGATGGCCCTGTACGACGTTGTTAGCAAGCTCCCGCTGGCGCTGA 8160
QY 8158 TGGCTCTCTCATACGATTTCAATTAATCTCCCAAGCAGCGGTGAGTTCCTGTTGAATA 8217
Db 8161 TGGAGACTCTCTACGATTCGAATTAATCACTACAGGACAGCGGTTGAATTCCTGTTGCAAG 8220
QY 8218 CTTGAAATCAAGAAATGCGCTATGGCTTCTCATATGACACCCGCTGTTTGTGCTCAA 8277
Db 8221 CGTGGAGTCCAAAGAAACCCGATGGGTTCTCGTATGATACCCGCTGTTTGTACTCCA 8280
QY 8278 CGGTCACTGAGAGTGAATCTGTTGAGGACTCAATTTACCAATGTTGTGACTGGCC 8337
Db 8281 CAGTCACTGAGAGCGACATCCGTAACGAGGAGGCAATTAACCAATGTTGTGACCTGGACC 8340
QY 8338 CCGAGGCGCAGACAGGCTAAGGTCGCTCACAGAGCGCTTTACATCGGGGTCCTCTGA 8397
Db 8341 CCCAAGCCCGGTGGCCATCAAGTCCCTCACTGAGAGGCTTATGTTGGGGGCCCTCTTA 8400
QY 8398 CTAACCTAAAAGGGCAGAACTGCGGTTATCGCGGTGCGCGCAAGTGGCGTGA 8457

Db 8401 CCAATTCRAAGGGGGAAAACTCGGCTACCGCAGGTGCGCGCAGCGCGTACTGACAA 8460
QY 8458 CTAGCTCGCGTAATACCCCTACATGTTACTTGAAGGCCACTCAGCCTCTCGAGCTGCAA 8517
Db 8461 CTAGCTGTGTAAACCCCTCACTTGTCTATCAAGGCCCGGACGCCCTGTGAGCGCGAG 8520
QY 8518 AGTCTCAGGACTGCACGATGCTCGTGAACGGAGACGACCTTGTGCTTATCTGTGAAGCG 8577
Db 8521 GGTCTCAGGACTGCACCATGCTCGTGTGGGCGAGCATTAGTCTGTGAAGTG 8580
QY 8578 CCGGAAACCCAGGAGGATCGGCGGCCCTACGAGCCTTCAAGAGGCTATGACTAGGTATT 8637
Db 8581 CGGGGTCCAGGAGGACGCGCGAGCCTCGAGAGCCTTCAAGAGGCTATGACCCAGGTACT 8640
QY 8638 CCGCCCCCGGGATCCGCCACCAAGATACGACCTCGAGCTGATAAATCATCATGTT 8697
Db 8641 CCGCCCCCGGGACCCCAACCAAGATACGACTTGGAGCTTATAACATCATGCT 8700
QY 8698 CTCCCAATGTGTGAGTGGGCGACGATGCTCTGGCAAAAGGATATCTACCTCACCCGTTG 8757
Db 8701 CCTTCAAGCTGTCAGTGGCGCGCTGGAAAGAGGCTTACTACTTACCTTACCCGTTG 8760
QY 8758 ACCCAACCAACCCCTTTCACGCGCTCGTGGGAGACAGCTAGACACATCCCAATCACT 8817
Db 8761 ACCCTAACCCCTTCGCGAGAGCGCGTGGGAGACAGACACATCCCAATCACT 8820
QY 8818 CTGGCTTAGCAATATCATGATGCGCCCAACCTATGGCAAGGATGATTCTGATGA 8877
Db 8821 CTTGGCTTAGCAATATGATGCTTGGCCCCACATGTTGGCGGAGGATGATCTGATGA 8880
QY 8878 CTCACTTTTCTCCATCTCTTCTAGCTCAAGAGCAATTTGAAAAAGCCCTGGATGTCAGA 8937
Db 8881 CCAATTTCTTTAGGCTCTCATAGCAGGATCAGCTTGAACAGGCTCTTAACTGTGAGA 8940
QY 8938 TCTACGGGCTTCTACTTCACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 8997
Db 8941 TCTACGGGCTTCTACTTCACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 9000
QY 8998 ATGCTCTTACGCAATTTACACTCCACAGTTTACTCTCCAGGTGAGATCAATAGGTTGCTT 9057
Db 9001 ATGCTCTTACGCAATTTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 9060
QY 9058 CATGCTCTCAGAAACTTTGGGCTACCCCTTGGCAACCTCGAGACATTCGGGCGCAGAAATG 9117
Db 9061 CATGCTCTCAGAAACTTTGGGCTACCCCTTGGCAACCTCGAGCTTGGAGACACCGGCGCGAGCG 9120
QY 9118 TCOCGCTTACGCTTCTGTCAGAGGAGGCGCGCTTCTGTCAGATACCTCTTTA 9177
Db 9121 TCOCGCTTACGCTTCTGTCAGAGGAGGCGCGCTTCTGTCAGATACCTCTTTA 9180
QY 9178 ACTGGGCGAGTAAGGACCAAGCTTAACTCACTCCCAATCCCGCGCGCTCCAGCTGGACT 9237
Db 9181 ACTGGGCGAGTAAGAACAAAGCTCAACTCACTCAATAGCGCGCTGGCGGCTGACT 9240
QY 9238 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9297
Db 9241 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9300
QY 9298 CCGGACCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9357
Db 9301 CCGGACCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9360
QY 9358 TGTCTCCCAACCGGATGAACCGGAGCTAAACCACTCCAGGC-CTTAAAGCCATTTCCCTGTT 9416
Db 9361 TCTTCCCAACCGATGAAGTTGGGTTAAACACTCCGCGCTCTTTAGGCCATTTCCCTGTT 9420
QY 9417 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9476
Db 9421 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9480
QY 9477 TTTCTTTT-----TCC 9487
Db 9481 TTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9540

QY	9488	CTTCTTTAATCGTGGCTGCATCTTTAGCCCTAGCTACCGCTAGCTGTGTAAGAGGTCCTGTGAG	9547
Db	9541	CTTCTTTAATGTTGGGTCCATCTTTAGCCCTAGCTACCGCTAGCTGTGTAAGAGGTCCTGTGAG	9600
QY	9548	CCGCATGACTCGCAGAGAGTGTCTGATATGCGCCTCTCTGCAGATCATCT	9595
Db	9601	CCGCATGACTCGCAGAGATCTGATATCTGCGCCTCTCTGCAGATCATCT	9648

RESULT 5

US-09-995-937-5
; Sequence 5, Application US/09995937
; Publication No. US20030028010A1
; GENERAL INFORMATION:

Query Match	65.3%	Score 6264;	DB 10;	Length 12980;
Best Local Similarity	78.4%;	Pred. No. 0;		
Matches 7565;	Conservative	0;	Mismatches 2030;	Indels 53;
				Gaps 3

QY	61	TCTTACG	CAGAAAGCGT	TACGCAT	TGGCGT	TAGTAT	TAGT	TAGT	GTCTG	CTG	CGTCC	AGGAC	120
Db	61	TCTTACG	CAGAAAGCGT	TACGCAT	TGGCGT	TAGTAT	TAGT	TAGT	GTCTG	CTG	CGTCC	AGGAC	120
QY	121	CCCCCTC	CCGGGAGAC	CCATAGT	GTCTG	CGGAA	CCGGT	TGAT	GATAC	CCGGAAT	TCCCA	180	
Db	121	CCCCCTC	CCGGGAGAC	CCATAGT	GTCTG	CGGAA	CCGGT	TGAT	GATAC	CCGGAAT	TCCCA	180	

181	QY	GACGACGGGTCCCTTTCTTTGGATCAACCCGCTCAATGCTCGAGATTTTGGGCTGCCCC	240
181	DB	GACGACGGGTCCCTTTCTTTGGATAAACCCGCTCAATGCTCGAGATTTTGGGCTGCCCC	240
241	QY	CGGAGACTGTAGCGGAGTAGTGTGGTTCGCGAAAGCCCTTGTGCTACTGCTGATAGG	300
241	DB	GCAAGACTGTCTAGCGGAGTAGTGTGGTTCGCGAAGGCTTTGTGTACTGCTGATAGG	300
301	QY	GTGCTTGCAGTGCCTCCGGGAGTCTCTGTAGACCGTGACCAATGAGCAGAACTCTAAAC	360
301	DB	GTGCTTGCAGTGCCTCCGGGAGTCTCTGTAGACCGTGACCAATGAGCAGAACTCTAAAC	360
361	QY	CTCAAGAAAACCAAAGTATACCAACACCGCCGCCACAGACCTCAAGTTCCTCGGGCG	420
361	DB	CTCAAGAAAACCAAAGTATACCAACACCGTCGCCACAGACCTCAAGTTCCTCGGGTG	420
421	QY	GTGGTCAGATCGTTGCTGTGAGTTTACCTGTTGCCCGCAGGGGCCCTCAGTTGGGTGTC	480
421	DB	GCGGTCAAGATCGTTGCTGTGAGTTTACTGTTGCCCGCAGGGGCCCTAGATTGGGTGTC	480
481	QY	GCGGACTAGGAAGCTTCCGAGCGGTTCGAACTCTGTGGAAAGGCGAACAATATCCCAA	540
481	DB	GCGGACGAGGAAGACTTCCGAGCGGTTCGAACTCGAGGTAGACGTCAACCTATCCCA	540
541	QY	AGGCTTCGCCACCGCAGGCGCTGGGCTCAGCCCGGGTACCTTCGCGCCCTCTATG	600
541	DB	AGGCACTGTCGCCCGCAGGCGCAGCACTTGGGCTCAGCCCGGGTACCTTCGCGCCCTATG	600
601	QY	GCNATGAGCGCTTGGGTGGCAGATGGCTCTCTGTCAACCCCGCGCTCCCGCTAGTT	660
601	DB	GCAATGAGGCTTGGGTGGCGGGAGTGGCTCTGTCTCTCCCGTGCTCTCGGCTAGCT	660
661	QY	GGGGCCCCGACCGACCCCGCGCTAGTTCGCGTAATTGGGTAAAGTCAATGATACCTTA	720
661	DB	GGGGCCCCACAGACCCCGCGCTAGTTCGCGCAATTGGGTAAAGTCAATGATACCTTA	720
721	QY	CATGGGCTTCCGCGATCTATGGGGTACATTCGCTCGTCGCGCGCCCTCAGGGGCG	780
721	DB	CGTGCGGCTTCGCGCACTCATGGGTACATACCGCTCGTCGCGCGCCCTCTTGGAGCG	780
781	QY	CTGCGAGGCTTGGCACAGGTCGCGGTTCTGGAGCAGCGCTGAATATGCAACAG	840
781	DB	CTGCCAGGGCCCTTGGCGCATGGGCTCGGGTTCTGGAAGCGCGTGAATATGCAACAG	840
841	QY	GGAACTTCCCGGTTCTTCTCTATCTTCTCTATGCTCTGTGCTCTGTCTGTGTGACCA	900
841	DB	GGAACTTCTGTGTGCTCTTCTCTATCTTCTCTGTGCGCTCTCTCTGTGCTGACCG	900
901	QY	TCCAGCTTCGGTTATGAAGTCGGCACTGTCCGGGATATACCATGTCAACGAACACT	960
901	DB	TGCCCCGTTTCAGGCTTACCAAGTCGCAATTCCTCGGGCTTTACCATGTCAACGAATGTT	960
961	QY	GCTCCAACTCAAGCATTGTATGAGSCAGGACGTGATCATGCACTATCCGGGTGCG	1020
961	DB	GCCCTAACTCGATTATGTGTATGAGSCGCGCATGCACTCTGTGCACTCCGGGTGTG	1020
1021	QY	TGCCCTGTGTTTCAGGAGGGTAACAGCTTCCGTTCTGTGGTAGCGCTCACTCCCAACGCTG	1080
1021	DB	TCCCTTCGCTTCGAGAGGTAACCGCTCGAGGTGTTGGGTGGCGGTCAACCCCAACGCTG	1080
1081	QY	CGGCGAGGATGCGAGCGTCCCACTACGACATACGACGCCAGTGCATTCGCTCGCTTG	1140
1081	DB	CCACCGGACGGCAAACTCCCAACACGAGCTTCGACGTCAATCGATCTGCTTCTGTCG	1140
1141	QY	GGACGGTGTCTTCTGCTCGCTATGTACGTGGGGATCTCTCGGATCTATTTCTCTCG	1200
1141	DB	GGACGGCAACCTCTGCTCGGCCCTTACGTTGGGGACCTGTGCGGGTCTGTCTTTCTTG	1200
1201	QY	TCTCCCAGCTGTTCACTTCTGCGCTCGCGCATGAGACAGTGCAGACTGCAACTGCT	1260
1201	DB	TTGCTACACTGTTTACCTTCTCTCCAGSCGCACTGACGACGAAGACTGCAATGCT	1260
1261	QY	CAATCTATCCGGCCATGTATCAGGTCACCGCATGGCTGGGATATGATATGAACTGGT	1320

1261 CTATCTATCCCGCCATPATACGGGTATCECATGGCATGGATATGATGATGAACTGGT 1320
 1321 CACCTACAAACAGCCCTAGTGGTGTGCGAGTGTCTCCGATCCACAAAGCTGTCTGTGACAA 1380
 1321 CCCCTACGGCAGCGTGTGGTGTAGTCTAGTGTCTCCGATCCACAAAGCCATCATGGACA 1380
 1381 TGTGGCGGGGGCCCACTGGGGAGTCTGGGGGCTTGGCTACTATCTCATGTTAGGGA 1440
 1381 TGATCGCTGTGTCTCACTGGGGAGTCTGGGGGCTAGGTATTTCTCATGTTGGGA 1440
 1441 ACTGGCTAAGTGTCTGATTTGGCGCTACTCTTTGCGGCTTGAAGGGGAGACCCACA 1500
 1441 ACTGGCGAAGTCTCTGT 1500
 1501 CGACGGGAGGGTGGCGGCGACACCACTCTGGGTTCACTGCTCTTTCTCATCTGGG 1560
 1501 TCACGGGGGAAGTGGCGGCGCACCACTGGCTGGGCTGTGTGTGTGTGTGTGTGTGTGT 1560
 1561 CGTCTCAGAAATCCAGCTTGTGATACCAACGCGAGTGGCAGATCAACAGGACTGCC 1620
 1561 CCAAGCAGAAACCACTGATCAACACCAACGCGAGTGGCAGATCAATAGACAGGCT 1620
 1621 TAAATTCGAATCACTCTCTCCCAACTGGGTTCTTTGCGGCTGTGTGTGTGTGTGTGTGT 1680
 1621 TGAATCGAATGAAGCTTACACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1680
 1681 TCAACTGTCTCGGGTGGCGGAGCGCATGGCGAGCTGGCGGCGCATGACTGTGTGTGGCC 1740
 1681 TCAACTCTTCAAGTGTCTGAGAGTGGCGAGTGGCGGCGCATGACTGTGTGTGTGTGTGT 1740
 1741 AGGGTGGGGCCCATCACTATTAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 1800
 1741 AGGGTGGGGTCTATCACTGT 1800
 1801 GGCATTAACCGCTCGACCTGT 1860
 1801 GGCATTAACCGCTCGACCTGT 1860
 1861 ATTGTTTCAACCGCTGT 1920
 1861 ATTGTTTCAACCGCTGT 1920
 1921 ATAGCTGGGGAGATCAGACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
 1921 ATAGCTGGGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 1981 GCAATGTTTCTGGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
 1981 GCAATGTTTCTGGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
 2041 CCCGTGTAAATCGGGGGGTGCTTAACGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2100
 2041 CCCGTGTGTATCGGGGGGTGCTTAACGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2100
 2101 GGAAGCACCGCGAGTCTTACAAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2160
 2101 GCAAGCATCCGAGGACCACTCTCTGGTGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 2160
 2161 GCTAGTAGACTACCATACAGCTTTGCACTTACCCCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220
 2161 GCAATGTTTCTGGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 2221 TTAAGTTAGGATGTATGTGGGGGTGGAGACAGGCTCAATCGCGCATGCAATTTGA 2280
 2221 TCAAGTCAAGGATGTATGTGGGGGTGGAGACAGGCTGGAGCGGCTGCAATTTGA 2280
 2281 CTCGAGGAGCGCTGTATCTTGGAGACAGGATAGGTGAGAACTCAGCCCGCTCTGCTG 2340
 2281 CGGGGGCGAACCGCTGT 2340
 2341 TGTCTCAACAGAGTGGCAGATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2400

2341 TGTCCACCAACAGTGGCAGGTCTTCTCGTGTCTTTCACGACCCCTGCCAGCCTTGTCCA 2400
 2401 CTGGTTTGTATCTTCATCCATCAGAACATCGTGCAATACCTGTAGTACGGTGTAGGT 2460
 2401 CGGCTCTATCACTCCACAGACATTTGTGACGTGAGTACTTGTACGGGTAGGT 2460
 2461 CAGCGTTTCTCTCTTTGCAATCAATGGAGTACATCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2520
 2461 CAGCATCGCTCTCTGGCCATTAAGTGGAGTACGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520
 2521 CAGACGGCGGT 2580
 2521 CAGACGGCGGT 2580
 2581 CTTTAGAGAACTTGT 2640
 2581 CTTTGGAGAACTTGT 2640
 2641 CTTTCTTGT 2700
 2641 CTTTCTCTGT 2700
 2701 CGTATGCTTTTATGGGTATGGCGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
 2701 TCTACGCTTCTACGGGATGTGCT 2760
 2761 CTTACGCTTGGACCGGGAGATGCTGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2820
 2761 CATACGCTTGGACCGGGAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2820
 2821 TATTTCTTGT 2880
 2821 TGGCGCTGACTGT 2880
 2881 AATTTCTTATCACCAGACCGGAGCGCATGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2940
 2881 AGTATTTTGTACCAAGT 2940
 2941 GGGGAGCGCGAGT 3000
 2941 GGGGGGGGGGATGCGGT 3000
 3001 ACATCAACAAACTCTGT 3060
 3001 ACATCAACAAACTCTGT 3060
 3061 CGAGAGT 3120
 3061 TTAAGT 3120
 3121 AAGT 3180
 3121 AGATAGCGGAGGT 3180
 3181 CGTACGTTTATACCACTTACCCACTGT 3240
 3181 CCTATGTGTATACCACTTACCCACTGT 3240
 3241 TTGCTGT 3300
 3241 TGGCGGT 3300
 3301 GAGCAGACCGCTGT 3360
 3301 GGGCAGATACCGCGGT 3360
 3361 GGAAGGAGATTTTGT 3420
 3361 GCCAGGAGATTTTGT 3420
 3421 CGCCCATCAGCGCTTCTCCCAACAAACGCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3480
 3421 CGCCCATCAGCGCTTCTCCCAACAAACGCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3480

7858 AGCCCTGCAAGTCAGCGCCGCCACATTCGGCCAAATCCAAATTTGGCTATGGGGCAAGG 7917
Db |
7861 AAGCTTGCAGCCTGAGCGCCGCCACATTCAGCCAAATCCAAATTTGGCTATGGGGCAAG 7920
Qy |
7918 ACCTCGGAAACCTTATCCAGCAGGCGCGTTAAACACATCCGCTCCGCTGGGGAGACTTGC 7977
Db |
7921 ACCTCGGTGGCCATGCCAGAGGCGGTAGCCACATCAACTCCGCTGGAAGACCTTC 7980
Qy |
7978 TGAAGACACTGAAAACACCAATTGACACCACTCATATGGCAAAAGTAGGTTTCTGGG 8037
Db |
7981 TGAAGACACTGAAAACACCAATTGACACCACTCATATGGCAAAAGTAGGTTTCTGGG 8040
Qy |
8038 TCCAAACAGAGAGGAGGCGCGCAAGCAGCTCGCTTATCGTATTCACAGACTCGGAG 8097
Db |
8041 TTGAGCTCTGAGAGGAGGCGGTGTAAGCCAGCTCGTCTATCGTGTTCGCCAGCTGGCG 8100
Qy |
8098 TTGCTGTATCCGAGAGATGCGCTTTTACGAGCTGCTTCCACCTTCCCTCAGGCGGTGA 8157
Db |
8101 TGGCGGTGCGAGAGATGCGCTTTTACGAGCTGCTTCCAAAGCTCCCGCTGGCGGTGA 8160
Qy |
8158 TGGGCTCCTCATAGGATTTCAATATCTCCCGCCAGAGCGGGTCTGAGTTCTCTGTGAATA 8217
Db |
8161 TGGGAGCTCCTACGGATTCCTAATCTACACAGGACAGCGGGTGAATTCCTCTGTGAAG 8220
Qy |
8218 CCTGGAAATCAAGAAATGCCCTATGGCTTCTCAATATGACACCGCTGTTTGAATCAA 8277
Db |
8221 CGTGGAAATCAAGAAATGCCCTATGGCTTCTCAATATGACACCGCTGTTTGAATCAA 8280
Qy |
8278 CGGTCACTGAGAGTGAATTCGTGTGAGAGTCAATTTACCAATTTGAGTGTGAGTGGCC 8337
Db |
8281 CAGTCACTGAGAGGACATCGTACGAGGAGGCAATTTACCAATTTGAGTGTGAGTGGAC 8340
Qy |
8338 CCGAGGCCAGAGGCCATGAGTTCGCTCAGAGCGGCTTACATCGGGGTCCCGTGA 8397
Db |
8341 CCCAAGCCCGCTGCGCCATCAAGTCCCTCACTGAGAGGCTTATGTTGGGGCGCCCTCTTA 8400
Qy |
8398 CTAACTCAAGAGGCGAGAACTGCGGTTATCCCGGTGCGCGCAAGTGGGCTGTGACGA 8457
Db |
8401 CCAATTCAGGGGGGAAACATGCGGCTACCGCAGGTGCGCGCGAGCGGCTACTGACAA 8460
Qy |
8458 CTAGCTGGGTAAATACCTCATCATGTTACTTGAAGGCGACATGAGCGCTGTGAGCTGCA 8517
Db |
8461 CTAGCTGGGTAAATACCTCATCATGTTACTTGAAGGCGACATGAGCGCTGTGAGCTGCA 8520
Qy |
8518 AGCTCCAGGATGTCAGCATGCTGCTGACGAGAGCAGCTTGTGCTATCTGTGCAAGCG 8577
Db |
8521 GGCTCCAGGACTGCACCATGCTCGTGTGTGGCGAGCACTTATGCTGTGAAAGTG 8580
Qy |
8578 CGGMAACCCAGGAGGATGCGGCGGCCCTACGAGCCTTCCAGGAGCTATGACTAGGTAT 8637
Db |
8581 CGGGGTCCAGAGGACGCGCGAGCCTTACGAGCCTTCCAGGAGCTATGACAGGTACT 8640
Qy |
8638 CCGGCCCGCCGGGATCCGCGCCCAACAGAAATGAGCAGCTGGAGCTGATACATCATGTT 8697
Db |
8641 CCGGCCCGCCGGGATCCGCGCCCAACAGAAATGAGCAGCTGGAGCTTAAACATCATGCT 8700
Qy |
8698 CCTCAATGTGTGAGTCCGCGCAGATGATCTGCGCAAGGATGATACCTTCAACCGCTG 8757
Db |
8701 CCTCAAGCTGTGAGTCCGCGCAGCAGCGCTTGGAAAGAGGCTTACTACCTTACCCCTG 8760
Qy |
8758 ACCCCACCAACCCCTTGCAGGGCTGGGTGGGAGACAGCTAGACACACTCCAATCAACT 8817
Db |
8761 ACCCTTACAACCCCTTGCAGAGCGCGGTGGGAGACAGCAACACTCCAGTCAAT 8820
Qy |
8818 CTGCGTAGGCAATATCATATGATGAGCGCCACCCCTTGGCAAGGATGATCTGATGA 8877
Db |
8821 CCGGTAGGCAACATAATCATGTTTGGCCCCACACTGTGGCGGAGGATGATGATGA 8880
Qy |
8878 CTCATTTTCTCCATCTTCTAGCTCAGAGCACTTGAAGAAACCTGGATCTGAGA 8937
Db |
8881 CCCATTTCTTAGCGCTCTCATAGCAGGATCAGCTTGAACAGGCTTAACTGTGAGA 8940

8938 TCTACGGGCTGTGCTACTCCATTTAGCCACTTGACCTACCTCAGATCATTTGACGACTCC 8997
Db |
8941 TCTACGGGCTGTGCTACTCCATTTAGCCACTTGACCTACCTCAGATCATTTGACGACTCC 9000
Qy |
8998 ATGCTCTTAGGGCAATTTACACTCCACAGTTACTCTCCAGGTGAGATCAATAGGGTGGCTT 9057
Db |
9001 ATGCTCTTAGGGCAATTTACACTCCACAGTTACTCTCCAGGTGAGATCAATAGGGTGGCG 9060
Qy |
9058 CATGCTCAGGAAACTTTGGGTTACACCCCTTCCGAACCTGGAGACATCGGCGCAGAAATG 9117
Db |
9061 CATGCTCAGGAAACTTTGGGTTACACCCCTTCCGAACCTGGAGACATCGGCGCAGAAATG 9120
Qy |
9118 TCCGCGCTAAGCTACTGTCCAGGGGGAGGCGCGCACTTGTGGCAGATACCTCTTGA 9177
Db |
9121 TCCGCGCTAAGGCTTCTGTCCAGAGGAGCGGCTGCCATATGTGGCAGATACCTCTTGA 9180
Qy |
9178 ACTGGCAGTAAAGCAAGCTTAAACTCACTCCAAATCCCGGCCCGCTCCAGCTGGACT 9237
Db |
9181 ACTGGCAGTAAAGCAAGCTTAAACTCACTCCAAATCCCGGCCCGCTCCAGCTGGACT 9240
Qy |
9238 TGTCTGGCTGTGCTCGCTGCTTACAGCGGGGAGACATATCACAGCCGTGTCTCGTG 9297
Db |
9241 TGTCCGCTTGTTCAGCGCTGCTACAGCGGGGAGACATATCACAGCCGTGTCTCGTG 9300
Qy |
9298 CCGGACCCCGCTGGTTCCGTTGTGCTTACTCTTCTTCTAGGGTAGCATTTACC 9357
Db |
9301 CCGGCGCCGCTGGTTCTGGTTTCTGCTACTCTGCTCGCTCAGGGGTAGCATTTACC 9360
Qy |
9358 TGTCTCCCAACCACTGAGCGGGAGTAACTCCAGGC-CTTAAAGCAATTTCTGTGTT 9416
Db |
9361 TCTCTCCCAACCACTGAGGTGGGTAAACACTCGGCTCTTAGGCCATTTCTGTGTT 9420
Qy |
9417 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9476
Db |
9421 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9480
Qy |
9477 TTTCTTTT-----TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9487
Db |
9481 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9540
Qy |
9488 CTCTTTTATGTTGGCTCCATCTTAGCTAGTCAAGGTAGCTGTGAAAGTCCGTGAG 9547
Db |
9541 CTCTTTTATGTTGGCTCCATCTTAGCTAGTCAAGGTAGCTGTGAAAGTCCGTGAG 9600
Qy |
9548 CCGCATGACTGCAGAGTCTCATAGTGGCTTCTGTGAGATCATGT 9595
Db |
9601 CCGCATGACTGCAGAGTCTCATAGTGGCTTCTGTGAGATCATGT 9648

RESULT 6
US-09-917-563-5
; Sequence 5, Application US/09917563
; Publication No. US20030073080A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAPERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBERS: US/09/917,563
; FILING DATE: 27-Jul-2001

```

; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/238,076
; FILING DATE: 26-JAN-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-917-563-5

Query Match 65.3%; Score 6264; DB 10; Length 12980;
Best Local Similarity 78.4%; Pred. No. 0;
Matches 7565; Conservative 0; Mismatches 2030; Indels 53; Gaps 3;

Qy 1 GCCAGCCCCCTGATGGGGGACACTCCACCATGATCACTCCCTCTGAGAACTACTG 60
Db 1 GCCAGCCCCCTGATGGGGGACACTCCACCATGATCACTCCCTCTGAGAACTACTG 60

Qy 61 TCTTCACGACGAAACGCTAGCCATGCGGTAGTATGAGTGTGCTGAGCTCCAGAC 120
Db 61 TCTTCACGACGAAACGCTAGCCATGCGGTAGTATGAGTGTGCTGAGCTCCAGAC 120

Qy 121 CCCCCCTCCGGGAGAGCCATAGTGGTCTGCGAAACCGGTGAGTACACCGAAATGCCAG 180
Db 121 CCCCCCTCCGGGAGAGCCATAGTGGTCTGCGAAACCGGTGAGTACACCGAAATGCCAG 180

Qy 181 GACGACCGGGTCTTTCTTGGATCAACCGCTCAATGCTGAGATTTGGGGTGCCTCC 240
Db 181 GACGACCGGGTCTTTCTTGGATCAACCGCTCAATGCTGAGATTTGGGGTGCCTCC 240

Qy 241 GCGACATGCTACCGAGTCTTGGTTCGGAAGCCCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 241 GCAAGACTGCTAGCGAGTGTGGTTCGGAAGCCCTGCTGCTGCTGCTGCTGCTGCTG 300

Qy 301 GTGCTTGCAGTGCCTCCGGGAGGTCTCTGATAGCCGTGACCATGAGCAGCAATCCTA 360
Db 301 GTGCTTGCAGTGCCTCCGGGAGGTCTCTGATAGCCGTGACCATGAGCAGCAATCCTA 360

Qy 361 CTCAAAGAAAACCAACCTAACAACCGCCGCCACAGGACGTCAGTTCCCGGGC 420
Db 361 CTCAAAGAAAACCAACCTAACAACCGCTCGCCACAGGACGTCAGTTCCCGGGT 420

Qy 421 GTGGTCAGATCGTGGTGGAGTTACTCTGTCGCGAGGCGCCAGCTGGGGTGGTGC 480
Db 421 GCGGTTCAGATCGTGGTGGAGTTACTCTGTCGCGAGGCGCCAGCTGGGGTGGTGC 480

Qy 481 GGGCGACTAGGAAGGTTCCGAGCGGTGCAACCTCGTGGAAAGGCGACACCTATCCCA 540
Db 481 GGGCGAGAGGAGACTTCCGAGCGGTGCAACCTCGTGGAGTAGAGCTAGCCTATCCCA 540

Qy 541 AGGCTCGCGACCCGAGGGGAGGGCTGGGCTCAGCCCGGTTACCTTGGCCCTCTATG 600
Db 541 AGGCACTCGCGCCGAGGGGAGGACCTGGGCTCAGCCCGGGTACCTTGGCCCTCTATG 600

Qy 601 GCATGAGGGCTGGGTGGGAGGATGCTCTGTCACCCCGCTCCCGGCTCCCGGCTAGT 660
Db 601 GCATGAGGGTTCGGGTGGGAGGATGCTCTGTCACCCCGCTCCCGGCTCCCGGCTAGT 660

Qy 661 GGGGCCCCCAGGACCCCGGCGTAGTTCGGTAAGTTCATCGATACCCCTTA 720

```

```

Db 661 GGGGCCCCCAGGACCCCGGCGTAGTTCGGTAAGTTCATCGATACCCCTTA 720
Qy 721 CATGGCGCTTCGCGGATCTCATGGGGTACATTCGGCTCGTCGGCGCCCCCTTAGGGGCG 780
Db 721 CGTGGCGCTTCGCGGACCTCATGGGTACATCCGCTCGTCGGCGCCCCCTTAGGGGCG 780
Qy 781 CTCGACAGGGCTTGGCACACAGGTGTCCGGTTCGAGGACGGCGTGAATGAAACAG 840
Db 781 CTCGACAGGGCTTGGCGCATGGGTCCGGTTCGGAAGACGGCGTGAATGAAACAG 840
Qy 841 GGAACCTTCGCGGTGCTCTCTATCTTCCTGCGGCTGCTCTCTGCTGCTGCTGCTG 900
Db 841 GGAACCTTCGCGGTGCTCTCTATCTTCCTGCGGCTGCTCTCTGCTGCTGCTGCTG 900
Qy 901 TCCAGCTTCGCTATGAAGTGCAGAACGTGTCGGGATATACCATGTCCAGAACGACT 960
Db 901 TCCAGCTTCGCTATGAAGTGCAGAACGTGTCGGGATATACCATGTCCAGAACGACT 960
Qy 961 GCTCCAACTCAAGCATTTGTATGAGGACGAGCGATGATCATGATCATCTCCGGGTGG 1020
Db 961 GCTCCAACTCAAGCATTTGTATGAGGACGAGCGATGATCATGATCATCTCCGGGTGG 1020
Qy 1021 TGCCCTGTGTCAGGAGGTAAACAGCTCCGCTGCTGGGTAGCGCTCACTCCAGCGCTG 1080
Db 1021 TGCCCTGTGTCAGGAGGTAAACAGCTCCGCTGCTGGGTAGCGCTCACTCCAGCGCTG 1080
Qy 1081 CGGCGAGGAATGCCAGCTCCCACTACGACAAATACAGCGCACGCTCGATCTGCTCTTG 1140
Db 1081 CGGCGAGGAATGCCAGCTCCCACTACGACAAATACAGCGCACGCTCGATCTGCTCTTG 1140
Qy 1141 GAGCGGTGCTTCTGCTCCGCTATGATGAGTGGGGATCTCTGGGATCTATTTTCTCTG 1200
Db 1141 GAGCGGTGCTTCTGCTCCGCTATGATGAGTGGGGATCTCTGGGATCTATTTTCTCTG 1200
Qy 1201 TCTCCAGCTGTTCACCTTCTCGCTCGCGGATGAGACGTCAGGACGTCGAACTGCT 1260
Db 1201 TCTCCAGCTGTTCACCTTCTCGCTCGCGGATGAGACGTCAGGACGTCGAACTGCT 1260
Qy 1261 CAATCTATCCGGCGATGATTCAGTTCAGCTCAGCTGGGTGGGATATGATGAACTGGT 1320
Db 1261 CAATCTATCCGGCGATGATTCAGTTCAGCTCAGCTGGGTGGGATATGATGAACTGGT 1320
Qy 1321 CACCTAACAAGCCCTAGTGTGCGAGTTGCTCCGGATCCCAAGCTGCTGCTGAGCA 1380
Db 1321 CACCTAACAAGCCCTAGTGTGCGAGTTGCTCCGGATCCCAAGCTGCTGCTGAGCA 1380
Qy 1381 TGGTGGGGGGCCCACTGGGGAGTCTGCGGCGCTGCGGCTGCTGCTGCTGCTGCTG 1440
Db 1381 TGGTGGGGGGCCCACTGGGGAGTCTGCGGCGCTGCGGCTGCTGCTGCTGCTGCTG 1440
Qy 1441 ACTGGGCTAAGGTTCTGATTTGCGCTACCTTTTGGCGGCTGCAAGGGGAGACCA 1500
Db 1441 ACTGGGCTAAGGTTCTGATTTGCGCTACCTTTTGGCGGCTGCAAGGGGAGACCA 1500
Qy 1501 CGACGGGAGGGTGGCGGCGACACCACTCCGGTTCAGTCCCTTTTCTGATCTGGG 1560
Db 1501 CGACGGGAGGGTGGCGGCGACACCACTCCGGTTCAGTCCCTTTTCTGATCTGGG 1560
Qy 1561 GGTCTCAGAAATCCAGCTTGTGATACAAAGGAGCTGGGACATCAACAGGAGTGGC 1620
Db 1561 GGTCTCAGAAATCCAGCTTGTGATACAAAGGAGCTGGGACATCAACAGGAGTGGC 1620
Qy 1621 TAAATTTGAAATGACTCCCTCCAACTGGGTTCTTTGGCGGCTGTTTACGACACA 1680
Db 1621 TAAATTTGAAATGACTCCCTCCAACTGGGTTCTTTGGCGGCTGTTTACGACACA 1680
Qy 1681 TCAACTCGTCCGGGTCCCGGAGCGCATGGCGAGCTGGCGGCGCATTTGAGTGGT 1740
Db 1681 TCAACTCGTCCGGGTCCCGGAGCGCATGGCGAGCTGGCGGCGCATTTGAGTGGT 1740
Qy 1741 AGGGTGGGGGCGGCTATCACTATATCAAGCTTAAACAGCTCGGATCAGAGGCTTATGCT 1800
Db 1741 AGGGTGGGGGCGGCTATCACTATATCAAGCTTAAACAGCTCGGATCAGAGGCTTATGCT 1800

```

Qy	1801	GGCAATTACGGCCTCGACCGTGTGGTGTGGTACCGCGCTCGAGGTGTGGTCCAGTGT	1860
Db	1801	GGCACTACCCCTCAAGACCTTGTGGCAATTGTGCCCGCAAGAGCGTGTGTGCCCGGTAT	1860
Qy	1861	ATTGTTTCACCCAAGCCCTGTGTGGTGGGACCAACCGATCGTTCGGGTGTCCCTAGCT	1920
Db	1861	ATTGCTTCACTCCAGCCCGTGTGTGGTGGGAACGACCGACAGGTGGGCGCGCTACCT	1920
Qy	1921	ATAGTCGGGGAGANTGACAGACGTGATGTCTCTCAACACAGCGCTCGGCCACAG	1980
Db	1921	ACAGTCGGGTGCAAAATGATACGGATGTCTTCGTCTCTTAAACACACAGCGCCACCGCTGG	1980
Qy	1981	GCACTGTTCCGCTGTATCATGGATGAATAGTACTGGTTTCACTAAGACGTGGGAGTTC	2040
Db	1981	GCAATTGTTGGTTGTATCTTGGATGAACTCACTGGATTCACCAAGTGTGGGAGCGC	2040
Qy	2041	CCCCGTGTAAATCCGGGGGTGGTTAAACCGCAACCTTGATTCGCCCAACGGACTGCTTC	2100
Db	2041	CCCTTGTGTTCATCGAGGGGTGGCAACAACACCTTGCTCTCGCCCACTGATGTGTTCC	2100
Qy	2101	GGAGCACCCCGAGGCTACTTACACAANVTGCTCTGGSGCCCTGGTTGACACCTTAGCT	2160
Db	2101	GCAAGCATCCGGAAGCCACATATCTCTGGTGGGCTCCGTTCCGTGATTCACCCAGGT	2160
Qy	2161	GCCTAGTAGATACCCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTCCATCT	2220
Db	2161	GCATGTCGACTACCCGTATAGGCTTTGGCACTATCCTTGTAACCATCAATTACACCATAT	2220
Qy	2221	TTAAGTTAGATGATGATGTGGGGGGGTGGAGCACAGGCTCAATGCCGATGCAANTGGA	2280
Db	2221	TCAAAGTCAGGATGTACGTGGAGGGGTTCAGACACAGGCTGGAGGGGCTGCAACTGGA	2280
Qy	2281	CTCGAGGAGAGCGCTGAATCTTGAGAGACAGGGATAGTCAAGAACTCAGCCCGCTGTGC	2340
Db	2281	CGCGGGCGAACCCTGTGATCTTGAAAGACAGGGACAGGTCCGAGCTCAGGCCATGCTGC	2340
Qy	2341	TGCTTACAACAGAGTGGCAGATCTGCCCTGTGTTTCAACACCTTACCGGCTTTATCCA	2400
Db	2341	TGTCACACACACAGTGGCAGGTTCCTTCGGTGTCTTTCAAGCCCTGCCAGCCTTGTTCCA	2400
Qy	2401	CTGGTTTGATCGATCTCCATCAGAACATCTGGAGCGTGCATATCTCTPACGGTGTAGGT	2460
Db	2401	CCGGCCTCATCCACTCCACAGAACATGTGGACGTGCATCTTGTACGGGGTAGGT	2460
Qy	2461	CAGCGTTTGTCTCTTTGCAATCAAAATGGAGTACATCTGTTCCTTTCTCTCTCTCTGG	2520
Db	2461	CAAGCATCGGTCCTGGGCCATTAAGTGGAGTACGTGTTCTCTCTCTCTCTCTGTTG	2520
Qy	2521	CAGACGGCGCGTGTGTGCTGTGCTGTGGATGATGCTGCTGATAGCCAGGCTAGACCG	2580
Db	2521	CAGACGGCGCGCTGCTCTCTCTGTGGATGATGTTACTCATATCCCAAGCGAGCGGG	2580
Qy	2581	CCATTAGAAACTTGTGTGTCCTCAATGGGGCGTCCGTGGCCGAGCGCATGTAATTTCT	2640
Db	2581	CTTTGGAAACCTCGTAACTCAATGACAGCATCCCTGGCCGGACGCACTCTGTGT	2640
Qy	2641	CTTTCTTGTTCTCTGCGCCGCTGTGATTAATTAAGGGCAGGCTGGCTCCTGGGGGG	2700
Db	2641	CTTCTCTGTGTTCTTCTGCTTGGGTGGTATCTGAAGGGTAGTGGGTGCCCCGAGCGG	2700
Qy	2701	CGTATGCTTTTATGGCGTATGCGCTGCTCTCTCTCTACTGSCGTTTACCAACAGAG	2760
Db	2701	TCTACGCCCTTACGGGATGTGGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2760
Qy	2761	CTTACGCCCTTGGACCGGGAGATGGGTGCATCGTGGGGGGTGGCTTCTTGTAGTCTGG	2820
Db	2761	CATAAGCATGGACACGAGGTGGCCGCTCGTGTGGGGCGTGTGTCTTGTGGGGTTAA	2820
Qy	2821	TATCTTGACCTTGTCAACATACTACAAGTGTTCCTACTAGGCCATATATGGTGTTAC	2880
Db	2821	TGGCGGTGATCTGTGTCGCATATTAACAAGCGTATCAGCTGGTGCATGTGGTGGCTTC	2880

QY	2881	AA	TACTTTAT	CACAGACGAGCGAGGCGCCACATGCAAGTGTGGGTCCCCCCCCTCAACGTTTC	2941
Db	2881	AG	TATTTTCTG	ACCAAGAGTAGAAGCGCAACTGCACGTGTGGGTTCCCCCCTCAACGTTCC	2941
QY	2941	GGG	AGGCGCGCATGCGCATCATCTCTCCTCACGTTGCGGTTCATCCAGAGTTAAATTTTG	3000	
Db	2941	GGG	GGGGCGGATCCGGTCATCTTACTCATGATGTGTGTTGTACACCGACTCTGGTATTTG	3000	
QY	3001	ACAT	CACCAAAATCTCTGTCTCGGCATATCGGCCCGCTCATGTGTCTCAGGTGSCATAA	3060	
Db	3001	ACAT	CACCAAAATCATCTCTGGGCATCTTGGACCCCTTTGGATTCTTCAAGCAGATTTCG	3060	
QY	3061	CGAG	GTGCCGTACTCTGTGCGGCTCAAGGGCTCATTCGTGCAATGATTTAGTTCGGA	3120	
Db	3061	TTAA	AGTCCCTACTCTGTGCGGTCTCAGGCTCTCCCGAATCTCCGCGCTAGCGCGGA	3120	
QY	3121	AAG	TCGCGGGGTCAATATGTCCAAATGGTCTTTCATGAAGCTGGCGGCTGACAGGTA	3180	
Db	3121	AGA	TACCGGAGTCAATACGTGCAAAATGCCATCATCAAGTTTATGGGCGCTTACTGGCA	3180	
QY	3181	CGT	ACGTTTATAACCATCTTACCCCACTCGGSACTGGGCCCAACGGGCTTACAGAAC	3240	
Db	3181	CCT	ATGTTATTAACCATCTCAACCTCTTTCGAGACTGGGCGCAACAGGCTTCGAGATC	3240	
QY	3241	TT	CGGTGGCGGTAGAGCCGTGCTCTTCTCCGCAATGAGACCAAGTTCATCACTGG	3300	
Db	3241	TG	CCGTGGCTGTGGAAACAGTCTGTCTCTCCCGAATGGAGACCAAGTTCATCACTGG	3300	
QY	3301	GAC	GACACCGCTGGTGTGGGACATCATCTTGGTCTTACCGCTCTCGCCGCGCAAGG	3360	
Db	3301	GGC	GAGATACCGCGCGTGGGTGTACATCATCAACGGCTTGCCGTCTCTGCGCGTAGGG	3360	
QY	3361	GG	AGGAGATATTTTGGACCGGCTGATAGTCTCGAAGGCGAAGGTGGCGACTCTTTG	3420	
Db	3361	GC	AGAGTACTTCTTGGCGCAGCGCATCATCAAGGATGTCTCCAAGGCTGGAGTTGCTGG	3420	
QY	3421	CG	CCATTCAGGCGCTACTTCCCAACAAAGCGGGCGTACTTGTGTTCATCATCACTAGCC	3480	
Db	3421	CG	CCATTCAGGCGCTACTTCCCAACAAAGCGGAGGCTCTCTTAGGCTTATAATCACCAGCC	3480	
QY	3481	TC	CAGCGCGGACAGAACCGTTCGAGGGAGGTCAAGTGTCTTACCGCAACAC	3540	
Db	3481	TG	ACTGCGCGGACAAACAAAGTGAGGTGAGGTCCAGATCTGTGTCAACTGTCTACCC	3540	
QY	3541	AA	CTTTCTCTGGCACTGTCATCAAGCGCGTGTCTGGACTGTCTACCATGGCGCTGGCT	3600	
Db	3541	AA	CTTTCTCTGGCACTGTCATCAAGCGGTGTCTGGACTGTCTACCAACGGGCGCGGAA	3600	
QY	3601	CGA	AGACCTTAGCGGTCCAAAAGTTCATTCACCAAAATGTACACCAATGTAGACCTGG	3660	
Db	3601	CGA	AGACCTTAGCGGTCCAAAAGTTCATTCACCAATGTAGACCTGTAGACCAAG	3660	
QY	3661	AC	CTCTGCTGCGTGCAGCGCGCCCCCGGGCGGCTCATGACACCAATGAGCTGTGGCA	3720	
Db	3661	AC	CTTGTGGGCTGCGCGGCTCTCAAGTTTCCGCTCATTTGACACCTTGACCTTCGGCT	3720	
QY	3721	G	CTCGACCTTTACTTTGGTCAAGAGCATGTGTGATGTCAATTCGGTGGCGCGGAGCG	3780	
Db	3721	C	CTCGACCTTTACTTTGGTCAAGAGCATGTGTGATGTCAATTCGGTGGCGCGGAGGTG	3780	
QY	3781	AC	AGCGGGAGGTACTCTTCCCGCTAGGCGGCTCTCTTCTTCAAGAGCTCTCTCGGTTG	3840	
Db	3781	AT	AGCGGGTAGCTGCTTTTCGCCCCGGCCCATTTCTACTTGAAGGCTCTCTCGGGG	3840	
QY	3841	G	CCAATTGCTTTGCCCTTCGGGACAGTGTGGGCGTCTTCCGGCTGTCTGTGCAACC	3900	
Db	3841	G	TCGCTTTGTGCCCGCGGACACCGCGTGGGCTTATTCAGGCGCGGCTGTGCAACC	3900	
QY	3901	GG	GGGCTCGGAAGGCGGTGACTTTCATACCCGTTGATGTCTATATGGAACTACCATGCGGT	3960	
Db	3901	GT	GGAGTGGCTTAAGGCGGTGACTTTCATACCCGTTGATGTCTATATGGAACTACCATGAGAT	3960	
QY	3961	CT	CCGCTCTTACAGACAACTCAACCCCCCGGCTGTATCCGAGACATTCAGTGGCAC	4020	

Db 3361 CCCCGGTTTCCAGGACAACTCTCTCCACAGAGTCCCCAGAGCTTCCAGGTGGCCC 4020
 QY 4021 ATCTGCAAGCTCTTCTGAGCGGCAAGAGCAACAAAGTCCCGGTGCGTATCGAGCCC 4080
 Db 4021 ACCTGATGCTCCACAGCGGAGCGGTAAAGACCAAGAGTCCCGGTGCGTATCGAGCCC 4080
 QY 4081 AAGGTTACAGGCTGCTGCTGTAACCGTTCGTTGCGGCACTTAGGTTTGGGGGT 4140
 Db 4081 AGGCTTCAAGGTGTTGGTGTCAACCCCTCTGTTGCTGCAACGCTGGGCTTGGTCTT 4140
 QY 4141 ATATGTTCAAGGCAACAGGTATCGACCTTAACATCAGAACTGGGGTAAGGACCAATACCA 4200
 Db 4141 ACAATGTCAGAGCCATCGGGTGTATCTTAATATCAGAGCCGGGTGAGAACAAATTACCA 4200
 QY 4201 CGGGCGGCTCATATACGATCTCCACTATGGCAAGTTCCTTGGCGACGGTGGCTGTTCTG 4260
 Db 4201 CTGGCAGCCCATACGTAATCCACTACGGCAAGTTCCTTGGCGACGGGCTGCTCAG 4260
 QY 4261 GGGCGGCTATCAGATCATATATATGATGATGCACTCAACTGACTCGACTCAATCT 4320
 Db 4261 GAGGTGCTTATGACATAATATATTTGTGACGAGTGCCACTCCAGGATGCCAATCCATCT 4320
 QY 4321 TGGGATCGGCACTCTCTGACCAAGCGGAGAGCGCTGAGCGCGCTCGTCTGCTGCTG 4380
 Db 4321 TGGGATCGGCACTCTCTGACCAAGCGGAGAGCGCTGAGCGCGCTCGTCTGCTGCTG 4380
 QY 4381 CCACGCTACACTCGGGATCGGTATCGGTGCGACACCCCAATATCAGAGAAATAGGCC 4440
 Db 4381 CCACTGCTACCCCTCGGGCTCGCTCACTGTCTCCCTAACATCGAGGAGGTGCTC 4440
 QY 4441 TGTCCAAAGTGGAGAGATCCCTCTATGGCAAGCCATCCCATTTGAGGCCATCAAG 4500
 Db 4441 TGTCCAAAGTGGAGAGATCCCTCTATGGCAAGCCATCCCATTTGAGGCCATCAAG 4500
 QY 4501 GGGGAGGATCTCATTTTCTGCCATCCAAAGAAATGTGACAGCTTCGCGCAAGC 4560
 Db 4501 GGGGAGGATCTCATTTTCTGCCATCCAAAGAAATGTGACAGCTTCGCGCAAGC 4560
 QY 4561 TGACAGGCTCGGATCGGATGAGCGTGTAGCATATTTACGGGGCTTGTGTCGCTCATAC 4620
 Db 4561 TGGTGGCATTTGGCATCAATGCGGTGGCTACTACCGGCTTTGACGCTGTGTCATCC 4620
 QY 4621 CGCCTATCGGAGCTGCTGCTGGGAGAGAGAGCTCTAATGACGGGTTTCAACGGG 4680
 Db 4621 CGACAGCGGCGATGTTGCTGCTGCTGCTGACCGATGCTCTAATGCTGCTTACCGG 4680
 QY 4681 ATTTGACTCAGTGTATCGATCAATATCATATGTTTCAACCGACAGTCTGCTGCTG 4740
 Db 4681 ACTTCGACTCTGTGATAGACTGCAACACGCTGTCTACTCAGACAGTCTGATTT 4740
 QY 4741 ATCCCACTTACCATTTGAGAGAGAGAGAGAGAGCTCTAATGACGGGTTTCAACGGG 4800
 Db 4741 ACCCTACCTTTTACCATTTGAGAGAGAGAGAGAGAGCTCTAATGACGGGTTTCAACGGG 4800
 QY 4801 GCGAGCTAGACTGAGAGAGGAGTGGGATCTACAGGTTTGTGATCTCAGAGAGAAC 4860
 Db 4801 GCGGAGAGGAGTGTGAGAGGAGAGAGAGAGAGAGCTCTAATGACGGGTTTCAACGGG 4860
 QY 4861 GGCCTTGGGCAATGTTGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
 Db 4861 GGCCTTGGGCAATGTTGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
 QY 4921 GGTATGAGCTCAGCCGCTGAGAGCTCGGTAGGTGAGGCTTACCTAAATACACAG 4980
 Db 4921 GGTATGAGCTCAGCCGCTGAGAGCTCAGTTAGGTGAGGCTTACCTAAATACACAG 4980
 QY 4981 GGTGCGGCTGCGAGGAGGAGTCTGAGTCTGAGGAGAGGCTTCTTACAGAGGCTCACCC 5040
 Db 4981 GGCTTCCGCTGCGAGGAGGAGTCTTGAATTTTGGGAGGAGGCTTCTTACAGGCTCACCTC 5040
 QY 5041 ACATAGATGCCCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5100

5041 ATATAGATGCCCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5100
 QY 5101 TGGCATATCAAGCTACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5160
 Db 5101 TAGGTTACCAAGGCAACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5160
 QY 5161 TGTGGAAGTGTCTATACGGCTGAAACCTTACATGCAACGGGCAACACCCCTCTCTATACA 5220
 Db 5161 TGTGGAAGTGTCTATACGGCTGAAACCTTACATGCAACGGGCAACACCCCTCTCTATACA 5220
 QY 5221 GGCTAGAGCGCTCCAAATGAGTCTATCTCACAACCCCAATACATTAATATCATATG 5280
 Db 5221 GACTGGCGCTGTCTCAGATGAAGTCACTGAGGAGCTGCTGAGGAGCTGCTGAGGAGCT 5280
 QY 5281 CATGATGTCGGTCACTGAGAGTCTCACTAGCACTGGTCTGCTGCTGCTGCTGCTGCTGCT 5340
 Db 5281 CATGATGTCGGTCACTGAGAGTCTCACTAGCACTGGTCTGCTGCTGCTGCTGCTGCTGCT 5340
 QY 5341 TTGCAAGCTTTGGCGGCAATATGCTGACGACAGGAGTGTGCTCATTTGTTGGGAGGATCA 5400
 Db 5341 TTGCAAGCTTTGGCGGCAATATGCTGACGACAGGAGTGTGCTCATTTGTTGGGAGGATCA 5400
 QY 5401 TCTTGTCCGGAGCGAGCTGCTGCTCCGACAGGAGAGTCTCTTACAGGAGTTCGATG 5460
 Db 5401 TCTTGTCCGGAGCGAGCTGCTGCTCCGACAGGAGAGTTCCTTACAGGAGTTCGATG 5460
 QY 5461 AGATGGAAGTGTGCTTCAACTTCTTACATCGAGAGGGAATCAGCTTCGCGGAGCTGCTG 5520
 Db 5461 AGATGGAAGTGTGCTTCAACTTCTTACATCGAGAGGGAATCAGCTTCGCGGAGCTGCTG 5520
 QY 5521 AATTCAAGCAAAAGGCGTCCGCTGTTGTTGCAAGCGGCACCAAGCAAGCGGAGGCTGCTG 5580
 Db 5521 AATTCAAGCAAAAGGCGTCCGCTGTTGTTGCAAGCGGCACCAAGCAAGCGGAGGCTGCTG 5580
 QY 5581 CTCCGCTGTTGAGTCCAAAGTGGGAGCGCTTGTGAGACCTTCTGGCGGAGCAGCATGTGGA 5640
 Db 5581 CTCCGCTGTTGAGTCCAAAGTGGGAGCGCTTGTGAGACCTTCTGGCGGAGCAGCATGTGGA 5640
 QY 5641 ATTTTCAATCAGCGGAATACAGTACCTAGCAGGCTTATCCACTCTGCTGGAACCCCGG 5700
 Db 5641 ATTTTCAATCAGCGGAATACAGTACCTAGCAGGCTTATCCACTCTGCTGGAACCCCGG 5700
 QY 5701 TAGCATATATGATGCAATTAACAGTCTTATACAGTCCGCTTACAGCAGGAGGAGGCTGCTG 5760
 Db 5701 TAGCATATATGATGCAATTAACAGTCTTATACAGTCCGCTTACAGCAGGAGGAGGCTGCTG 5760
 QY 5761 TCCGCTTTAAACATCTTGGGGGATGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5820
 Db 5761 TCCGCTTTAAACATCTTGGGGGATGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5820
 QY 5821 CAGCTTTTGGCGGCGGCGCATTCGCGGAGCGGCTGTTGGGAGCATATAGGCTTGGGAGG 5880
 Db 5821 CAGCTTTTGGCGGCGGCGCATTCGCGGAGCGGCTGTTGGGAGCATATAGGCTTGGGAGG 5880
 QY 5881 TCTGCTGGAATCTTGGGGGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5940
 Db 5881 TCTGCTGGAATCTTGGGGGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5940
 QY 5941 AGGTATGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6000
 Db 5941 AGGTATGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6000
 QY 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060
 Db 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060
 QY 6061 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6120
 Db 6061 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6120
 QY 6121 ACCAGCTCTCCCTAGCAGCTATGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6180
 Db 6121 ACCAGCTCTCCCTAGCAGCTATGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6180

6181 TCCTCTAGCTTACCATCACTCACTGCTGAGGGCTCCACAGTGGATTATAGG 6240 QY
6181 TACTCAGAGCTCTACTGTAACCCAGCTCTGAGGAGCTGCTAGTGGATAGCTCG 6240 Db
6241 ACTGCTCTACGCCATGCTCCGGCTCGTGGCTAAGGATGTTTGGATATGATACGG 6300 QY
6241 AGTTACCACTCACTGCTCGGTTCTGCTAAGGACATCTGGACTGGATATCGAG 6300 Db
6301 TGTGTGACTGTTAAGACTGGCTCCAGTCCAACTCTGCGGGTTACCGGAGTCC 6360 QY
6301 TGTGAGCGACTTTAAGACTGGCTGAAAGCCAAAGCTCATGCTCAAACTGCTGGATT 6360 Db
6361 CTTTCTGTCATGCCAAACCGGCTACAGGGAGTCTGGCGGGGACGCGCATCATGCAA 6420 QY
6361 CTTTGTGTCCTGCCAGCGGGTATAGGGGGTCTGGCGGAGACGCGCATATGCCA 6420 Db
6421 CCACCTGCCATTCGGGAGCAGATCCCGGACATGTCAAAACGGTTCTATGAGATCG 6480 QY
6421 CTCGCTGCCACTGTGAGCTGAGATCACTGGACATGTCAAAACGGGACGATGAGATCG 6480 Db
6481 TAGGGCTAGAACTCGAGAACACGTGGCACGGAGGTTCCCATCAACGATACCA 6540 QY
6481 TCGCTCTAGGACTCGAGAACATGTGGAGTGGAGCTTCCCATTAACGCTACCA 6540 Db
6541 CGGGACCTTCACACCTTCCCGCGGCCAACTATTCCAGGGCGCTATGGCGGTTGGCTG 6600 QY
6541 CGGGCCCTGTACTCCCTTCTTGGCGGAACTATAAGTTCCGCTGTGGAGGTTGCTG 6600 Db
6601 CTGAGAGTACGTGGAGTTACGCTGTGGGGATTTCCATACGTGACGCGCATACCA 6660 QY
6601 CAGAGGAATACGTGGAGTAAAGCGGTTGGGGACTTCCACTACGTATCGGGTATGACTA 6660 Db
6661 CTCACACGTAAAGTCCCATGCTCAGGTTCCGGCCCGCAATCTTCAACGAGTGGATG 6720 QY
6661 CTGACAACTTAAATGCCGTGCCAGATCCATCGCCCGAATTTTTCACAGAAATGAGC 6720 Db
6721 GAGTGGCGTTGCAAGTACGCTCCGGCTGCCAAACCTTTCTACGGGAGGAGCTCACGT 6780 QY
6721 GGTGGCCCTACATAGTTTGGCCCTTTCAGAGCCCTTGTCTGGGAGGAGGTATCAT 6780 Db
6781 TCAGAGTGGGCTGAAACATATCTGCTGGGTGGAGTCCCATGCGAGCCCGAACCG 6840 QY
6781 TCAGAGTGGACTCCACAGTACCGGTGGGTGGCAATACCTTGGAGCCCGAACCG 6840 Db
6841 ACCTAACAGTCTTACTTCCATGCTCAGCATCCCTCCCATATACAGCAGACAGCGCTA 6900 QY
6841 ACCTAGCGGTGTTGAGTCCATGCTACTGATCCCTCCCATATACAGCAGACAGCGCG 6900 Db
6901 AGCGTAGGCTGGCTAGAGGTTCTCCCGCTCTTTAGCCAGCTCATAGCTAGCCAGTTGT 6960 QY
6901 GGAGAGGTTGGCGAGAGGTTCAACCCCTTCTATGGCCAGCTCTCGGCCAGCCAGCTGT 6960 Db
6961 CTGGCCCTTTTGAAGGCGATGCACTACCCAGCTACCTCCCGAGCTCCCGAGCTCACTCA 7020 QY
6961 CCGCTCCATCTCTCAAGCACTTGCACCGCCACCATGATCTCCCTGACCCGAGCTCA 7020 Db
7021 TCGAGGCCAACCTCTTGTGGCGGAGAGATGGCGGAAACATCACTCGCTGGAGTCAG 7080 QY
7021 TAGAGGCTAACTCTGTGGAGGAGAGATGGCGGAAACATCACAGGTTGAGTCAG 7080 Db
7081 AGAATAAGGTAGTAATCTTGGACTTTTTCGAACCGCTTCAACCGGAGGGGATGAGCGG 7140 QY
7081 AGAACAAAGTGGTATTCTGGACTCTTCTGATCCGCTTGTGGCAGAGGAGGATGAGCGG 7140 Db
7141 AGATATCCGTCCGCGGAGATCTTGGAAATCCAGGAATTTCCCTCAGGTTGCCCA 7200 QY
7141 AGGTCTCCGCTACCCGCAAAATCTTGGGAAATCTCGGAGATTCGCCGCGCTTCCCG 7200 Db
7201 TATGGGACGCGGACTTACATCTCCACTGCTAGAGTCTTGGAGGACCGCGGACTACG 7260 QY
7201 TTTTGGCGCGGCGACTTACAAACCCCGCTAGTAGAGACGTTGGHAAAGCCTGACTACG 7260 Db

7261 TCCCTCCGGTGGTACACGATGCCATTTGCCACCTAACAAGGCTCTCTCCAAATCAACCTC 7320 QY
7261 AACCACTGTGTCTCATGGCTGCCGCTACCACTCCACGGTCCCTCTCTGCTCCGC 7320 Db
7321 CACGGAGAAAGAGACGTTGCTCTGACAGATCAATGTCTCTCTGCTCTGCTTGGCGGAGC 7380 QY
7321 CTCGGAATAAGCGTACGGTGTCTCTACCGATCAACCTATCTACTGCTTGGCCGAGC 7380 Db
7381 TCGGCATTAAGACCTTCGTTAGCTCCGATCTGTCGGCGTTGATAGCGGACCGGACCG 7440 QY
7381 TTGCCACAAAGTTTGGCAGCTCTCAACTTCGGCATTAACGGGCGACAAATACGACAA 7440 Db
7441 CCGTCTCTGACTGGCTCTCCAGACGGTGAACAGATTCGAGTTCGACTCTCTCT 7500 QY
7441 CATCTCTGAGCGCGCTCTCTGCTGCCCCCGACTCCGAGCTTCGAGTCTTATTTCT 7500 Db
7501 CCAATCCCGCTTGAAGGGAGCGGGGACCCGATCTCAGGACCGGGCTTTGGTCTA 7560 QY
7501 CCAATCCCGCTTGAAGGGAGCGCTGGGATTCGGATCTCAGCAGCGGTCATGCTCGA 7560 Db
7561 CGGT---GAGTGAGAGGCTAGTGAGGATGTGCTGTCTGCTCAATGCTCTATAGCTGA 7617 QY
7561 CGGTCACTAGTGGGCGCGACACGGAAGATCTGCTGTCTCTCATGCTCTTATTTCTGGA 7620 Db
7618 CAGGCGCTGATCAACGCTGCGGAGGAAAGTAAAGCTGCCATCAACCGTTGA 7677 QY
7621 CAGGCGCACTCTGCTACCCCGTGGCGGAAAGAAACAAACTGCGCATCAACGACTGA 7680 Db
7678 GCAACTCTTGTGCTGCTACCAACATGCTTACGCCCAACATCCCGCAGCGCAAGCC 7737 QY
7681 GCAACTCTGCTGCTACGCCATCACTGCTGCTTATTCACCACTTCAACGAGTCTGCT 7740 Db
7738 TCGGCGAAGAAAGTCACTTTGACAGATTCGCAAGTCTCTGATCATTTACCGGAGC 7797 QY
7741 AAAGGCAAGAAAGTCACTTTGACAGCTCAAGTTCTGACAGCCATTCAGGAGGAGC 7800 Db
7798 TACTCAGAGATGAGGCGGAGGCTCCACAGTTAAGCTTAAGCTTCTATCTATAGAG 7857 QY
7801 TECTCAAGAGTCAAGCAGCGGCTCAAAAGTGAAGCTTAACTTCTATCTCGTAGAG 7860 Db
7858 AGGCTTCAAGTCAAGCTTCCGCGCAATTCGCGCAATTCGCGCTATGCGGCAAGG 7917 QY
7861 AAGCTTCAAGCTTCAAGCTTCCGCGCAATTCGCGCAATTCGCGCTATGCGGCAAGG 7920 Db
7918 AAGCTTCAAGCTTCAAGCTTCCGCGCAATTCGCGCAATTCGCGCTATGCGGCAAGG 7977 QY
7921 AAGCTTCAAGCTTCAAGCTTCCGCGCAATTCGCGCAATTCGCGCTATGCGGCAAGG 7980 Db
7978 TGAAGACACTGAAACCACTTTGACACCACTCATGCGCAAAAGTGAAGTTTCTGCG 8037 QY
7981 TGAAGACACTGAAACCACTTTGACACCACTCATGCGCAAAAGTGAAGTTTCTGCG 8040 Db
8038 TCAACCCAGAGAGGAGGCGCGCAAGCTGCTTCTGCTTATTCGTAATTCGAGACCTGGAG 8097 QY
8041 TCAACCCAGAGAGGAGGCGCGCAAGCTGCTTCTGCTTATTCGTAATTCGAGACCTGGAG 8100 Db
8098 TCGTGTATTCGAGAGAGTGGCGCTTCAAGCTGCTTCAAGCTTCTGCTTATTCGTAATTCGAGACCTGGAG 8157 QY
8101 TCGCGGTGTCGAGAGATGGCGCTGACAGCTGCTTCAAGCTTCTGCTTATTCGTAATTCGAGACCTGGAG 8160 Db
8158 TGGGCTCTCTCATAGGATTTCAATATCTTCCCGCAAGCAGCGGCTCGAGTTCTGCTGAATA 8217 QY
8161 TGGGAGCTCTCTAGGATTTCAATATCTTCCAGCAGCAGCGGTTGAATTTCTGTCAG 8220 Db
8218 CCGTGAATCAAGAAATGCGCTTATGGCTTCTCAATGACACCGCTGTTGACTCAA 8277 QY
8221 CCGTGAATCAAGAAATGCGCTTATGGCTTCTCAATGACACCGCTGTTGACTCAA 8280 Db
8278 CCGTCACTGAGAGTCAATTCGTTGAGGAGTCAATTTACCAATGTTGACTTGGCC 8337 QY
8281 CAGTCACTGAGAGCAGATCCGTACGAGGAGGCAATTTACCAATGTTGACTTGGACC 8340 Db
8338 CCGAGGCCAGACAGGCCATAAGTCTGCTCAAGAGCGGCTTTACATCGGCGGCTCCCTGTA 8397 QY

2581 CTTGGAGAACCTCGTAAATACTCAATGACGATCCCTGGCCGGACCGACGGCTCTGTGT 2640 Db
2641 CTTTCCTGTGTCTCTTCTGCGCCGCTGTGTACATTAAGGCGAGCTGTGTCTCTGGGGGG 2700 QY
2641 CTTTCCTGTGTCTCTTCTGCGCCGCTGTGTACATTAAGGCGAGCTGTGTCTCTGGGGGG 2700 Db
2701 CGTATGCTTTTATGCGGTATGCGCGCTGTCTCTGCTCTTACCTGCGCTTACCAACACGAG 2760 QY
2701 TCTACGCTTCTACGGGATGTGGCTCTCTCTGCTCTCTGCTGCGCTTACCTGCGCTTAC 2760 Db
2761 CTTACGCTTCTGAGCGGAGATGCTCTGCTCTGCTGCGGGGTGCGCTTCTTGTAGTCTG 2820 QY
2761 CATACGCACTGGAACAACGAGGTGTCGCGCTGCTGCGGGGTGCTTCTTGTGCGGGTAA 2820 Db
2821 TATTCCTGACCTGTGTACCAATACAAAGTGTCTCTACCTAGGCTCATATGTTGGTTAC 2880 QY
2821 TGGCGCTGACTGTCTACCAATATACAGCGCTATATCAGCTGGTCAATGTGSGCTTC 2880 Db
2881 AATACTTTATCACAGAGCGAGCGCACATGCAAGTGTGGTCCCGCTCCCTCAACGCTTC 2940 QY
2881 AGTATTTCTGACACAGAGTAGAAGCGCACTGACGCTGTGGTTCCCGCTCTCAACGCTTC 2940 Db
2941 GGGAGGCGCGGATGCCATCATCTCTCAGTGTGGGTTCATCCAGAGTTAATTTTG 3000 QY
2941 GGGGGGGCGGATGCCCTATCTTACTCATGTGTGTACACCCGACTCTGGTATTTG 3000 Db
3001 ACATCACCAAACTCTGCTGCCATCTCGCGCCGCTCATGGTCTCCAGGCTGSCATAA 3060 QY
3001 ACATCACCAAACTCTGCTGCCATCTCGCGCCGCTCTTGGATTTCTTCAAGCCGAGTTGC 3060 Db
3061 CGAGAGTGCCTACTCTGTCGCGCTCAAGGCTCATCTGTCATGATGTTAGTGGAA 3120 QY
3061 TTAAGTCCCCACTCTGTCGCGCTTCAAGCCCTCTCCGATCTGCGCGTACGCGGA 3120 Db
3121 AAGTGGCGGGGTCAATATGTCCTCAAAATGCTTCAAGAAAGTGGCGGCTGTACAGTA 3180 QY
3121 AGATAGCGGAGGTCAATACGTCAAAATGCGCATCAAGTGTGGGGCGCTTACTGGCA 3180 Db
3181 CGTAGTTTAAACATCTTACCCACTGCGGACTGGGCCCGCCAGCGGGCTACAGACC 3240 QY
3181 CCTATGTATAAACATCTACCCCTCTTCGAGACTGGCGGCACAAACCGCTTCGAGATC 3240 Db
3241 TTGCGTGGCGGTAGAGCCGCTGCTCTCCGCTATGAGACCAAGGCTCATCACTGG 3300 QY
3241 TGGCGCTGGCTGTGGAACAGTCTCTTCTCCGNAATGAGACCAAGCTCATCACTGG 3300 Db
3301 GAGCAGACCCGCTGCTGTGGGACATCATCTTGGGTCTACCGCTCTCGCGCCGAGGG 3360 QY
3301 GGGCAGATACCGCGCTGCGGTGACATCATCAACGGCTTGCCTCTCTGCGCGTAGGG 3360 Db
3361 GGAAGGATATTTTGGAGCGGCTGATGCTTCGAGGCGAAGGTTGGGACTCCCTTG 3420 QY
3361 GCCAGGATATCTGTTGACACGCGACGGAATGCTCTCCAGGGGTGAGGTTGCTGG 3420 Db
3421 CGCCCATCAGCGCTTACTCCCAACAAACGCGGGGTACTTGGTTCATCATCACTAGCC 3480 QY
3421 CGCCCATCAGCGCTTACTCCCAACAAACGCGGGGTACTTGGTTCATCATCACTAGCC 3480 Db
3481 TCACAGCGCGGACAAAGAACAGGTGGAAGGAGGTTCAAGTGTCTTACCGCAACAC 3540 QY
3481 TGACTGGCGGACAAAACCAAGTGGAGGTGAGTCCAGATCGTGTCAACTGTCTACCC 3540 Db
3541 AATCTTCTGCGGACTGCATCAAGCGGTGTGCTGGCTGTCTTACCATGCGCTGGCT 3600 QY
3541 AAAACCTTCTGCGGACTGCATCAAGCGGTGTGCTGGCTGTCTTACCATGCGCTGGCT 3600 Db
3601 CGAAGACCTTAGCGCTTCAAAAGGTCCAAATCAACCAATGTACCAATGTAGACCTGG 3660 QY
3601 CGAGGACCATCGATCAACCAAGGTCTCTGTATCAAGTGTATACCATGTGAGACCAAG 3660 Db
3661 ACTCGTGGCTGGAGGCGCCCGCGGGCGGTCTCAATGACCAAGCACTGTGGCA 3720 QY
3661 ACTTGTGGCTGGCGCGCTCTCAAGGTTCGCGCTCATTTGACACCCCTGCACTGCGGT 3720 Db

3721 GCTCGACCTTTACTTGGTCAAGACATCTGATGTCTATTCGGTGCCTCGGCGAGGG 3780 QY
3721 CTTTCGACCTTTACTTGGTCAAGACATCTGATGTCTATTCGGTGCCTCGGCGAGGG 3780 Db
3781 ACAGCAGGGAAGTCTACTCTCCCGAGCGCTCTCTTACCTGAAAGGCTCTCGGTG 3840 QY
3781 ATAGCAGGGAAGTCTACTCTCCCGAGCGCTCTTACCTTACCTTAAAGGCTCTCGGTG 3840 Db
3841 GTCCATGTCTTTCCTCTCGGSGCAGTCTGCGGCTCTTTCGGGCTGTGTGTGACCC 3900 QY
3841 GTCCGCTGTGTGCTCCCGCGGACACCGCTGCGCTATTCAGGCGCGGTGTGACCC 3900 Db
3901 GGGCGTTCGGAAGGCGGTGGAATTCATACCCGTTCAGTCTATGGAATATACCATCGGT 3960 QY
3901 GTGAGTGGCCAAAGCGGTGGAATTCATACCCGTTCAGTCTATGGAATATACCATCGAT 3960 Db
3961 CTCGCTCTTCAGACATCAACCTCAACCCCGCTGTACCGACATTCACAGTGGCAC 4020 QY
3961 CCGCGTGTTCAGGACATCTCTCTCCACAGCTGCCCCAGAGCTTCCAGGTGGCCC 4020 Db
4021 ATCTGACGCTCTCTACTGCGAGCGCAAGAGCACCAAGTGCCTGCTGTATGCAAGCCC 4080 QY
4021 ACTGCTATGCTCCACCGGCGGTGTAGAGCACCAAGTCCCGCTGTGTATGCAAGCCC 4080 Db
4081 AAGGTAACAAGTGTCTCTCTGAAACCGCTTCTGCGGCACTTTCAGGCTTGGGGCT 4140 QY
4081 AGGCTACAAGTGTGTGCTCAACCCCTCTGCTGTGCAACGCTGGGCTTGGTCTT 4140 Db
4141 ATATGTCAGGACACAGCTATCGACCTTAACATCAGAACTGGGGTAAGACCATTACCA 4200 QY
4141 ACATGTCAGGCGCATGGGTGCTGATCTTAATATCAGGACCGGGGTGAGAACAAATACCA 4200 Db
4201 CGGCGCTCCATTAAGTCTCAACCTATGCGAAGTTCCTTTCGCGAGCTGGCTGTCTG 4260 QY
4201 CTGACGCGCATCACTGCTCAACCTATGCGAAGTTCCTTTCGCGAGCTGGCTGTCTG 4260 Db
4261 GGGCGCTATGACATCATATATGATGAGTGCCTCAACTGACTGCTGACATCCATCT 4320 QY
4261 GAGCGCTATGACATCATATATGATGAGTGCCTCAACTGACTGCTGACATCCATCT 4320 Db
4321 TGGCATGCGGACAGTCTCTGGAACAGGAGAGCGGTGAGGCGGCTGTGTGTGCTG 4380 QY
4321 TGGCATGCGGACAGTCTCTGGAACAGGAGAGCGGTGAGGCGGCTGTGTGTGCTG 4380 Db
4381 CCAGCTACACTCTCGGCTGCTTACCTGCGCACACCCCAATATCGAGGAATAGGCC 4440 QY
4381 CCAGCTACACTCTCGGCTGCTTACCTGCGCACACCCCAATATCGAGGAATAGGCC 4440 Db
4441 TGTCCAAATGAGAGATCCCTTCTATGCGAAAGGCTTCCCATTTGAGGCGCTTCAAG 4500 QY
4441 TGTCCAAACAGGAGATCCCTTCTATGCGAAAGGCTTCCCATTTGAGGCGCTTCAAG 4500 Db
4501 GGGGAGGACATCTCATTTCTGCGCATTCACAGAGAAATGAGAGCTCGCGCAAGC 4560 QY
4501 GGGGAGGACATCTCATTTCTGCGCATTCACAGAGAAATGAGAGCTCGCGCAAGC 4560 Db
4561 TGACAGGCTCGGACTGAAACGCTGTAGCATATTCAGGCGCTTGTGTGTCTCATAC 4620 QY
4561 TGTGCGCATTTGGGATCAATGCGCTGCGCTACTTACCGGAGTGTGACGCTGTCTATCC 4620 Db
4621 CGCCTATCGGAGAGTGTGTGTGTGTGCAACAGAGCTTCAATGACGCGTTTCAACGG 4680 QY
4621 CGACAGCGGAGTGTGTGTGTGTGTGCGACCGATGCTCTCATGACTGGCTTACCGCG 4680 Db
4681 ATTTGATCTGATGATGACATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4740 QY
4681 ACTTGCATCTGT 4740 Db
4741 ATCCACCTTCAACATTCAGAGAGCGACCTGTCGCCCAAGAGCGGCTGTGTGTGTGTGT 4800 QY
4741 ACCCTACCTTTACCATTTGAGACACCGCTTCCCGAGATGCTGTGTGTGTGTGTGTGTGT 4800 Db

6961 CCGCTCCATCTCTCAGACCAACTTGCACCGCCAAACCATGATCCCTCTGACCGGAGCTCA 7020
 7021 TCGAGGCAACCTCTTCTGGGGCAGAGATGGCGGAACATCACTCGGGTGAATCAG 7080
 7021 TAGAGGTAACCTCTCTGGAGGAGAGATGGCGGCAACATCAACAGGGTTGAGTCAG 7080
 7081 AGAATAAGGTAGTAATCTGAGACTCTTTTCGAACCGCTTCACGCGAGAGGGGATCAGAGG 7140
 7081 AGAACAAAGTGTGATCTGACTCTCTTCGATCCGCTGTGGCAGAGAGGATCAGCGG 7140
 7141 AGATATCGCTCGCGGAGAGATCTCTGGAAATTCAGGAAGTTCCTCTCAGCGTTGCCA 7200
 7141 AGGTCTCGTACCGCGCAAAATCTCGGAAATCTCGGAGATTCGCCCCAGCCCTGCCCC 7200
 7201 TATGGGCAACCGCGACTACAAATCTCTCACTGCTAGAGTCTGGAAGGACCGGACTACG 7260
 7201 TCTGGGCGCGCGGACTACAAACCCCGCGCTAGTAGACGTGGAAGGACCGGACTACG 7260
 7261 TCCCTCCGCTGTGTACACGGATGCCCCATTGCCACCTACCAAGGCTCTCTCAATACCACTC 7320
 7261 AACCACTGTGTTCATGGCTGCCCGCTACCACTCCACGCTCCCTCTCTGTGCTCCG 7320
 7321 CACGAGAAAGAGAGAGGTGTCTGTGACAGAAATCCAAATGCTGTCTCTGCTTGGCGAGC 7380
 7321 CTCGAAAAGAGGTACGGTGTCTCTCACGAATCAACCTATCTACTGCTTGGCGGAGC 7380
 7381 TCGCACTAAGACCTTCGTGTAGTCTCGGATCGTGGCGGTTGATAGCGGACCGGCAACG 7440
 7381 TTGCCAACAAAGTTTGGGAGCTCTCTCACTTCGCGGATACGGGCAACATACGACAA 7440
 7441 CCTTCTCTGACCTCCGCTCCGACGAGGTGACAAAGGATCCGACGTGAGTGTGTACTCT 7500
 7441 CATCTCTGAGCCCGCTCTCTGCTGCTCCCGCTCCGCTCCGCTGAGTCTCTCTCT 7500
 7501 CCAATCCCCCTTGAAGGGAGCGCGGAGCCCGATCTACGACAGGCTCTGCTGTA 7560
 7501 CCAATCCCCCTTGAAGGGAGCGCTGCGGATCTCGAGACGGGTCTGCTGTA 7560
 7561 CGGT---GAGTGAGGAGGTAGTGAGGATGCTGTGCTGTCAATGTCTCTATAGTGA 7617
 7561 CGGTGAGTGTGGGCGGACACGGAAGATGCTGTGCTGTCAATGTCTCTATGCTGGA 7620
 7618 CAGGCGCCCTGATCAGCCATCGCTGCGGAGGAAAGTAAGTGCCTCAACCCGCTGA 7677
 7621 CAGGCGCACTCTCTACCCCGCTGCGCTGGGAGGAAACAAATGCTCCCACTCAACGCACTGA 7680
 7678 GCACTCTTGTGTGTGCTACCAACATGCTGTACCCACACATCCCGAGGCAAGCC 7737
 7681 GCACTCTGTGTGTACCCCATCAATCTGTGTATTCACCACTTCACGAGTGTGCTG 7740
 7738 TCCGCGAAGAGGTACCTTTGACAGATGTCAAGTCTCGATGATCATTTACCGGAGC 7797
 7741 AAGGCGAAGAAAGTCACTTTGACAGACTGCAAGTCTGAGACGCCATACCAAGAGC 7800
 7798 TACTAAGAGATGAAGGCGAGGCTTCAAGTTAAGGCTAAGCTTCTATCTATAGAG 7857
 7801 TGCTAAGGAGTCAAGCAGCGCGCTCAAAAGTGAAGGCTAACTTGTCTATCCGTAGAG 7860
 7858 AGGCTGCGAGCTGACGCCCCCATCTTCGCGCAATCCAAATTTGCTATGGGCAAGG 7917
 7861 AAGCTTGCAGCTTGCAGCGCCCCCATCTTCAGCCAAATCCCAAGTTTGGTATGGGGCAAG 7920
 7918 ACCTCGGAACCTATCCAGCAGGCGGCTTAAACACATCCCGCTCGCTGTGGGAGGACTTC 7977
 7921 ACGTTCGTTGCCATGCGAAGAGGCGGTAGCCCACTCACTCCGTTGGAAGACCTTC 7980
 7978 TGGAGACACTGAACCAACCAATGACACCACTCATGATGGCAAAAGTGAAGTTTCTGG 8037
 7981 TGGAGACAGTGTAAACCAATAGACACTACCATCATGCGCAAGAACAGGTTTCTGCG 8040
 8038 TCCAAACAGAGAGGAGGCGGACGAGCTGCTTATCTGATTTCCCAAGACTGGAG 8097
 8041 TTCAGCTGAGAAAGGGGTCGTAAAGCCAGCTGTCTCATCTGTCTCCCGACTGGCG 8100

8098 TTCTCTATSCGAGAAAGATGGGCCCTTTTACGACGTGTCTCCACCCTTCTCAGGCGGTGA 8157
 8101 TGGCGCTGTGCGAAGAGATGGGCCCTGTACGACGTGTGTAGCAAGCTCCCCCTGGCGGTGA 8160
 8158 TGGGCTCTCTCATACGAGATTTCAATCTCCGCCAAAGCAGCGGCTCGAGTTCTCTGTGAATA 8217
 8161 TGGGAGCTCTCTACGGATCCCAATATCACAGGACAGCGGTTGAAATCTCTGTGCAAG 8220
 8218 CTTGGAATCAAAAGATGCTTATGGCTTCTCATATGACACCCGCTGTTTGTGACTCAA 8277
 8221 CTTGGAAGTCAAGAGAGACCCGATGGGTTCTCGTATGATACCCGCTGTTTGTGACTCCA 8280
 8278 CGGTCACTGAGATGACATTCGTGTGAGAGTCAATTTACCAATGTTGAGACTTGGCC 8337
 8281 CAGTCACTGAGACGACATCCGTACGGAGAGGCAATTTTACCAATGTTTGTGACTTGGACC 8340
 8338 CCGAGGCGACAGAGGCCATTAAGGTTCGCTCACAGAGCGGCTTTACATCGGGGCTCCCTGA 8397
 8341 CCCAGCCCGCTGGGCCATCAAGTCCCTCACTGAGAGGCTTTACGTTGGGGGCCCTCTTA 8400
 8398 CTAACTCAAAAGGCGAGAACTGCGGTTATCCCGGTTCGCGGAAAGTGGCGTGTGTGACGA 8457
 8401 CCAATTCAGGGGGGAAACTTGGGCTTACCGAGGTCCGCGGAGCGGCTTACTGACAA 8460
 8458 CTAGCTCGGTATATACCTTACATCTTACTTGAAGGCACTGACAGCCTGTGAGCTGCAA 8517
 8461 CTAGCTGGTTAAACACCTCTCTTGTACATCAAGAGCCCGGGAGGCTGTGAGCCGCGAG 8520
 8518 AGTCTCAGGACTGCAACGATGCTCTGTAACGAGACGACCTTGTCTGTATCTGTGAAAGCG 8577
 8521 GGTCTCAGGACTGCAACGATGCTCTGTGTGCGACGACTTAGTCTGTATCTGTGAAAGTG 8580
 8578 CCGGAAACCCAGGAGGATCGCGGCCCTTACGAGCTTACGAGGCTATGACTAGGTATT 8637
 8581 CCGGGGTCAGAGAGGAGCGCGGAGCCTGAGAGCCTTTACGAGGCTATGACAGGTACT 8640
 8638 CCGCCCCCGCGGAGTCCGCCCCCAACAGAAATAGCCTCGAGCTGATACATCATCTT 8697
 8641 CCGCCCCCGCGGAGCCCCCAACAGAAATAGCCTTGGAGCTTATACATCATCTT 8700
 8698 CCTCCAAATGTCTAGTTCGCGCACGATGCTCTGCGAAAGGATATACCTTCAACCCCTG 8757
 8701 CTTCCAAAGTGTAGTTCGCGCACGAGCGGCTTGAAGAAAGGCTTACTACTTACCTCCG 8760
 8758 ACCCCACACCCCTTTCGACGCGGTGCGTGGGAGACAGCTAGACACACTCAATCAACT 8817
 8761 ACCCTTACACCCCTTTCGCGGAGCGCGTGGGAGACAGCAAGACACACTCCAGTCAAT 8820
 8818 CTTGCTAGGCAAT 8877
 8821 CTTGCTAGGCAAT 8880
 8878 CTCATTTTCTCCATCTCTTAGCTCAAGAGCACTTGAAGAAAGCCCTTGAATTTCTGA 8937
 8881 CCCATTTCTTAGCTCTCTTAGCTCAAGAGCACTTGAAGAAAGCCCTTGAATTTCTGA 8940
 8938 TCTAGCGGCTTGTCTACTTCAATTTAGCCACTTGAACCTCAGATCATTTGACACTCC 8997
 8941 TCTAGCGGCTTGTCTACTTCAATTTAGCCACTTGAACCTCAGATCATTTGACACTCC 9000
 8998 ATGCTTCTAGCGCTTGTCTACTTCAATTTAGCCACTTGAACCTCAGATCATTTGACACTCC 9057
 9001 ATGCTTCTAGCGCTTGTCTACTTCAATTTAGCCACTTGAACCTCAGATCATTTGACACTCC 9060
 9058 CATGCTCAGGAAACTTTGGGTACCACTTCCGACCTGAGACATCGGGCCAGAAATG 9117
 9061 CATGCTCAGGAAACTTTGGGTACCACTTCCGACCTGAGACATCGGGCCAGAAATG 9120
 9118 TCCGCTTAAAGCTTGTCTCCAGGGGGGAGGCGCGCTTGTGAGGAGATACCTTCTTA 9177
 9121 TCCGCTTAAAGCTTGTCTCCAGGGGGGAGGCGCGCTTGTGAGGAGATACCTTCTTA 9180

QY 961 GCTCAACTCAAGCATTTGTATGAGGAGCGGACCTGATCATGATCACTTCCCGGTGG 1020
DB 961 GGCCTAACTCGAGTATTTGTACGAGGCGGCGATCCATCTCTGCACACTCCGGGTGTG 1020
QY 1021 TGCCCTGTGTTGAGAGGGTAAACAGTCCCGTGTGCTGGTAGCGCTCACTCCACGCTCG 1080
DB 1021 TCCCTTTGGGTTCCGAGGGTAAACGCTTCAGATGTTGGTGGCGGTGACCCCAAGGTG 1080
QY 1081 CGGCGAGAAATGCCAGCGTCCCACTACGACAAATACGACGCCACGCTCGACTTGTCTG 1140
DB 1081 CCACGAGGAGCGGAACCTCCCAACACGCACTTCGAGCTCATATCGATCTGCTTGTG 1140
QY 1141 GGACGCTGCTTTCTGCTCCGATATGTACGTGGGGATCTCTCGGATCTATTTTCCTCG 1200
DB 1141 GGAGCGCCACCTCTGCTCGGCCCTCTAGCTGGGGACCTGTGGCGGTCTGCTTTCTTG 1200
QY 1201 TCTCCAGCTGTTCACCTTCTCGCTCGCGGCGCATGAGACAGTGCAGGACTCGACTGCT 1260
DB 1201 TTGGTCAACTGTTTACCTTCTCTCCAGCGCCACTGGACGACGCAAGACTCGAATTGT 1260
QY 1261 CAATCTATCCCGCCATGTATCAGGTCACCGCATGGCTTGGGATATGATGATCAACTGCT 1320
DB 1261 CTATCTATCCCGCCATATAGGGTTCATCGCATGGCATGGATATGATGATGAACTGCT 1320
QY 1321 CACCTAACACGCCCTAGTGTGTGTCAGTGTCTCCGATCCCAAGCTGTGCTGTGGACA 1380
DB 1321 CCCCCTACGGCAGCGTGTGTGTAGCTCAGCTGCTCCGGATCCCAAGCCATCATGGACA 1380
QY 1381 TGSTGGCGGGGCCCACTGGGAGTCTTGCGGGCTTGCCTACTATTCATGTTAGGA 1440
DB 1381 TGATGCTGTGTCTCACTGGGAGTCTTGCGGGCATAGCTGATTTCTCCATGTTGGGA 1440
QY 1441 ACTGSECTAAGTCTGTGATGTGGGCTACTCTTTGCCGCGTGTGACGGGAGACCCACA 1500
DB 1441 ACTGGCGGAGTCTGT 1500
QY 1501 CGAAGGGAGGTGGCGGCCACACCACTCCGGGTTCACGTCCTTTCTCATCTGGGG 1560
DB 1501 TCACCGGGGAGTGGCGGCCACACCGGCTGGGCTTGTGTGTGTGTGTGTGTGTGTGTGT 1560
QY 1561 CGTCTCAGAAATCCAGTGTGATACCAACGCGAGCTGGACATCAACGAGCTGCC 1620
DB 1561 CGAAGCAGAACTCAAACTGATCAACCAACGCGAGTGGACATCAATATGACAGGCT 1620
QY 1621 TAAATGCAATGAATCCCTCCAACTGGGTTCTTTGCCGCGTGTGTTACGACACAACT 1680
DB 1621 TGAATGCAATGAAGCTTTACACCGGCTGTTAGCAGGGCTCTTCTATCAGCACAAT 1680
QY 1681 TCAATCTGTCGGGTGCCGAGCGCATGGCAGCTGCCGCCCAATGACTGGTTGGGCC 1740
DB 1681 TCAACTCTTTCAGGCTGTCTGAGAGGTTGGCAGCTGCCAGCGCTTACCGAATTTGCC 1740
QY 1741 AGGGTGGGGCCCATCACTATATCAAGCTTAAAGCTCGATCAGAGCGCTTATGCT 1800
DB 1741 AGGGCTGGGGTCTTACATGATTCAGTATTCAGAGAGCGGCTTCAGAGAACCCCTACTGCT 1800
QY 1801 GCAATACCGCTCGACCGT 1860
DB 1801 GCACTTACCTCCAGACCTTGT 1860
QY 1861 ATTGTTTACCCCAAGCTTGT 1920
DB 1861 ATTGCTTCACTCCAGCCCGGT 1920
QY 1921 ATAGCTGGGGGAGATGAGACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
DB 1921 ACAGCTGGGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980
QY 1981 GCACTGCTTCCGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
DB 1981 GCAATGCTTCCGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040

QY 2041 CCGGTGTAAATCGGGGGTTCGTTAAACCGCACTTGTATCTGCCCCACGAGCTGCTTCC 2100
DB 2041 CCGCTTGTGTATCGAGGGGTGGGCAACAACACTTGTCTCTGCCCACTGATTGTTCC 2100
QY 2101 GGAACACCCGAGGCTACTTACACAAAATGTGGCTCGGGGCCCTGTGTGACCTAGGT 2160
DB 2101 GCAAGCATCCGGAAGCCCACTACTCTGGTGGCTCCGCTCCCTGGATTAACCCAGGT 2160
QY 2161 GCCTAGTAGACTACCCATACAGGCTTGGCACTACCCCTGCACTCTCAATTTTCCATCT 2220
DB 2161 GCATGTTGCACTACCCGATATAGGCTTGGCACTATCTTGTACCACTCAATTACACCAAT 2220
QY 2221 TTAAGTTAGGATATATGTGGGGGCGTGGAGCAAGGCTCAATGCGCATGCAATTTGGA 2280
DB 2221 TCAAAAGTCAGGATGTACGTGGGAGCGGCTCGAGCAAGGCTGGAAGCGGCTCAACTGGA 2280
QY 2281 CTCGAGGAGCGCTGTAACTTGGAGGACAGGATAGGTCAGAACTCAGCCGCTGCTGC 2340
DB 2281 CGCGGGGCGAAGCTGTGATCTGGAAGACAGGACAGGTCGAGCTCAGCCCATTTGCTGC 2340
QY 2341 TGTCTAACACAGAGTGGCAGATACCTGCCCTGTGCTTCCACCACTTACCGCTTTATCCA 2400
DB 2341 TGTCCACACACAGTGGGAGGCTCTCGTGTCTTTTCAAGACCTTCCAGGCTTGTCCA 2400
QY 2401 CTGGTTTGAATCAATCTCATCAGAACATCTGTGGAGCTGCAATACCTGTACGGTGTAGGT 2460
DB 2401 CCGGCTCATCCACTCCACCAAGACATGTGGAGCTGCACTTGTACGGGTGTAGGT 2460
QY 2461 CAGCGTTTGTCTTCTTTCGATCAAAATGGGAGTACATCTGTGCTTTCCTTCTCCTGG 2520
DB 2461 CAAGCATCGCTCTTGGGCCATTAAGTGGGAGTACGTGCTTCTCTCTTCTTCTTCTGTTG 2520
QY 2521 CAGACCGCGGTGTGTGCTTGTGGATGATGCTCTCTGATAGCCAGGCTCAGGCGG 2580
DB 2521 CAGACCGCGGTGTGTGCTTGTGGATGATGCTCTCTGATAGCCAGGCTCAGGCGG 2580
QY 2581 CTTAGAGAACTTGGTGTCTCAATGCGGCTCGTGGCGGAGCGCATGATTTCTCT 2640
DB 2581 CTTTGGAGAACTCTGTAATCTCAATGAGCATCCCTGGCGGAGCGCATGCTTGTGT 2640
QY 2641 CTTTCTTGTGTTCTTCTGCGCGCTGTACATTAAGGCGAGGCTGCTCTCGGGCGG 2700
DB 2641 CTTTCTTGTGTTCTTCTGCGCGCTGTACATTAAGGCGAGGCTGCTCTCGGGCGG 2700
QY 2701 CGTATGCTTTTATGGCGTATGGCGGCTGCTCTCTCTTACTTGGCGTTTACACACAG 2760
DB 2701 TCTAGCGCTTCTACGGGATGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
QY 2761 CTTAGCGCTTGAACCGGAGATGGCTGCACTGTCGGGGGTGCGGTTCTTGTAGTCTGG 2820
DB 2761 CATAGCACTGGACACGAGGTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820
QY 2821 TATTTCTGACCTTGTCACTACTACAAAGTGTCTTCTACTAGGCTCATATGTTGTTTAC 2880
DB 2821 TGGCGCTGACTCTGTGCGCATTTACAGCGGTACATCAGCTGGTGCAATGTTGGCTTC 2880
QY 2881 AATATTTTATCACAGAGCGGAGCGCATGCAAGTGTGGGTCCCGGCTTCAAGCTTC 2940
DB 2881 AGTATTTTCTGACAGAGTGAAGCGCACTGCAAGTGTGGGTCCCGGCTTCAAGCTTC 2940
QY 2941 GGGGAGCGCGATGCCATCACTCTCTCTCACTGTCGGGGGTGCGGTTCTTGTAGTCTGG 3000
DB 2941 GGGGAGCGCGGATGCCATCACTCTCTCTCACTGTCGGGGGTGCGGTTCTTGTAGTCTGG 3000
QY 3001 ACATCACAACCTCTGCTCGCCATCTCGGCGCTCATGTTGCTTCCAGGCTGCGATTA 3060
DB 3001 ACATCACAACCTCTCTGCGCATTTCCGACCCCTTGGATTTCTTCAAGCCAGTTTGC 3060
QY 3061 CGAGAGTCCGCTTCTGCTGGCGCTCAAGGCTCATTCGTCGATGTCATGTTAGTGGCA 3120
DB 3061 TTAAGTCCCTTACTTCTGCGCGTTCAGGCTTCTCCGATCTGCGGCTAGCGGGA 3120
QY 3121 AAGTCGCGGGGTCAATTATGTCAAATGGTCTTCTCATGAAGCTGGGCGGCTGACAGGTA 3180

Db 3121 AGATAGCCGAGGAGTCAATACGTCGAATGGCCATCATCAAGTTAGGGCGCTTACTGGCA 3180
QY 3181 CGTACGTTTATACCAATCTTACCCCACTGGGGAAGCTGGGGCCACGGGGGCTTACGAGACC 3240
Db 3181 CCTATGTATACCACTCTCACCCCTCTTCGAGACTGGGCGACAAACGGCGCTGCGAGATC 3240
QY 3241 TTCCGGTGGGCTAGAGCCGCTGCTCTCTCCGCCATGGAGACCAAGGTCATCACTGGG 3300
Db 3241 TGGCCGTGGCTGAGAACAGAGTCTCTCTCCGAATGGAGACCAAGCTCATCACTGGG 3300
QY 3301 GAGCAGACACCGCTGGGTGGGACATCATCTTGGGTCTACCCGCTCTCCGCCAAGGG 3360
Db 3301 GGCAGATACCGCGCGTGGGTGACATCATCAAGGCTTGGCCGCTCTCCGCCGAGGG 3360
QY 3361 GGAAGAGATATTTTGGGACCGGTGATCTCGAAGGCGAAGGTCGACTCCTTG 3420
Db 3361 GCCAGGAGATACGCTTGGGCCAGCGCGAAGGATGCTCTCAAGGGGTGAGGTTGCTGG 3420
QY 3421 CGCCCATCACGGGCTACTCCCAACAAACGGGGGCTTGGTTCGATCATCACTAGCC 3480
Db 3421 CGCCCATCACGGGTACGCCACGACGAGAGAGGCTCTTAGGGTGATATATCACAGCC 3480
QY 3481 TCACAGCCCGGACAAAGAACCAAGGTGGAAGGGAGGTTCAAGTGGTTTACCGCAACAC 3540
Db 3481 TGACTGGCCGGGACAAACCAAGTGGAGGTGAGGTCCAGATCGTGTCAACTGCTACCC 3540
QY 3541 AATCTTTCCTGGGACCTGATCAACGGCGTGTCTGGAGTCTACCATGGCGCTGGCT 3600
Db 3541 AAACCTTCTTGGCAACGTGATCAATGGGGTATGCTGAGTGTCTACCGGGGCGGAA 3600
QY 3601 CGAAGACCTTAGCGGTCCAAAGGTCCAAATCACCCAAATGTACCAAAATGTAGACCTGG 3660
Db 3601 CGAGGACCATCGATCAACCAAGGCTCTGTATCCAGATGTATACCAATGTGACCAAG 3660
QY 3661 ACTCTGTGGTGGCGCGCTCCTCAAGGTTCCCGCTCATTEGACACCCCTGCACCTGGCGCT 3720
QY 3721 GCTCGGACCTTTACTTGGTACGAGACATGCTGATGTCTATCCGGTGGCGGGAGCGG 3780
Db 3721 CCTCGGACCTTTACTTGGTACGAGGACCGCGAATGTCATTCCTGCGCGCGGAGGTG 3780
QY 3781 ACACAGGGGAAGTCTACTCTCCCGGAGGCGGCTCTCTACTGAAAGGCTCTCCGGGTG 3840
Db 3781 ATAGCAGGGTAGCGTGTCTGCGCGGCAATTTCTTACTTGAAGGCTCTCTCGGGG 3840
QY 3841 GTCCATTCCTTGGCCCTTGGGGGACGTCGTGGGGGCTTCCGGGCTGCTGTGTGACCC 3900
Db 3841 GTCCGCTGTGTGCGCGCGGGAACGCGCGTGGGCTTATTCAGGGCGCGGCTGTGACCC 3900
QY 3901 GGGGGGTGCGAAGCGGTGACTTCTATACCCGTTGAGTCTATGGAACCTACCATGCGGT 3960
Db 3901 GTGAGTGGCTTAAGCGGTGAGCTTTATTCCTGTGGAGAACTTAGAGAACCAATGAGAT 3960
QY 3961 CTCGGGTCTTACAGACAACTCAACCCCGCGGTGTACCGACAGACATTCOAAGTGGAC 4020
Db 3961 CCCCCTGTTTCAAGCAAACTCTCTCCACAGCAGTGCCTCCAGAGCTTCCAGGTGGCC 4020
QY 4021 ATCTGACGCTCTTACTGGGACGGGACAGACCAAGTGGCGGCTGCTGTATGACGCC 4080
Db 4021 ACCTGCATGCTCCACCGGACGGGTAAGACCAAGGTCCCGGCTCGTACGACGCC 4080
QY 4081 AAGGTTACAGGTGCTGCTGTAACCGCTCGTTGGCGGCACTTAGGGTTTGGGGCGT 4140
Db 4081 AGGGCTACAGGTGTGTGTGCTCAACCTCTGTGTGCTGCAACGCTGGGCTTGGTGT 4140
QY 4141 ATATGTCCAGGACACCGGTATCGACCTTAACATCAGAACTGGGTAAGGACCAATACCA 4200
Db 4141 ACATGTCCAGGCCCATGGGTGATCTCTTAATCAGGACCGGGTGAGAACCAATACCA 4200
QY 4201 CGGGGGCTCTATAGTACTCCACCTATGGCAAGTCTTCTTGGCCGACGGTGGCTTCTG 4260

4201 CTGGCAGCCCATCATCGTACTCCACTACGGCAAGTTCCTTGGCCACGGCGGGTCTCAG 4260
QY 4261 GGGGGCGCTATGACATCATATAATATGTGATGAGTGCACATCAACTGACTCGACTACATCT 4320
Db 4261 GAGGTGCTTATGACATATATATTTGTGAGAGTGCACATCCACGGATGCGACATCACTCT 4320
QY 4321 TGGGCATCGGCACAGTCTCTGGACCAAGCGGAGACGGCTGGAGGGCGGTCTGTGTGCTCG 4380
Db 4321 TGGGCATCGGCACAGTCTCTTGACCAAGCAGACAGTGGGGGGCGAGAGCTGGTTGTGCTCG 4380
QY 4381 CCAACGCTACACCTCGGGATCGGTTACCGTGCCACACCCCAATATCAGAGAAATAGGCC 4440
Db 4381 CCACTGCTACCCCTCCGGGCTCCGTCAGTGTGCCATCTCCATCAACATCAGAGAGTTGCTC 4440
QY 4441 TCTCCAAACAATGAGAGATCCCTTCTATGGCAAGCCATCCCATTTAGGSCCATCAAGG 4500
Db 4441 TGTCCACACCGGAGAGATCCCTTTTACGGCAAGGCTATCCCTCGAGGGTGATCAAGG 4500
QY 4501 GGGGAGGCAATCTCATTTTCTGCCATTTCCAGAGAAATGTGACGAGCTCGCGCAAGC 4560
Db 4501 GGGGAAGACATCTCATCTTCTGCCACTCAAAGAGAAGTGCAGCAGGCTCGCGGGAAGC 4560
QY 4561 TGACAGGCTTCGACATGACGCTGTAGCATATTTACCGGGGCTTGTATGTCTCGTCATAC 4620
Db 4561 TGTCTGCAATGGGCATCAATGCGGTGGCTACTACCGCGTCTTGAAGTGTCTGTCTC 4620
QY 4621 CGCCTATCGGAGACGTCGTTGTCTGGCAACAGACGCTCTAATGACGGGTTTACCGGG 4680
Db 4621 CGACAGCGCGGATGTTGCTGCTGTGCGACCGATGCTCTCATGACTGGCTTTACCGGG 4680
QY 4681 ATTTGACATGAGTATCGATCGCATATCATGTGTACCCAGACAGTCTGATCTGAGCTTGG 4740
Db 4681 ACTTGCATCTGTGATGACTGTCAACACGCTGTCTCATCAGACGTCGATTTGAGGCTTGG 4740
QY 4741 ATCCCACTCTCCATTTGAGAGACGACCGCTGCCCAAGACGGGCTGTCGCTGCTCAAC 4800
Db 4741 ACCTTACCTTTACCATTTGAGACCAACCGCTCCCGAGATGCTGTCTCCAGGACTCAAC 4800
QY 4801 GGCAGAGTAACTGGCAGGGGTAGAGTGGCATTTACAGTTTGTGATTCACAGAGAAC 4860
Db 4801 GCGGGGCGAGGACTGCGAGGGGAAGCCAGGATCTACAGATTTGTGGACCGGGGAGC 4860
QY 4861 GGCCTCGGGCATGTTTCGATCTTCGCTCTGTGAGTGTCTATGACGCGGCTGTGCTT 4920
Db 4861 GCGCTCTCGGCATGTTTCGATCTTCGCTCTGTGAGTGTCTATGACGCGGGGTGTGCTT 4920
QY 4921 GGTATGAGTCAAGCGCGCTGACACCTCGGTTAGGTTGGGGCTTACCTAAATACACAG 4980
Db 4921 GGTATGAGTCAAGCGCGCGAGACTACAGTTAGGCTACGAGCGTACATGAACACCGCG 4980
QY 4981 GGTTCGCGCTCTCCAGGACCATCTCGAGTTCTGGAGAGCGTCTTACAGGCTCACCC 5040
Db 4981 GGTTCGCGCTCTCCAGGACCATCTTGAATTTTGGAGGGGCTCTTTACGGGCTCACCTC 5040
QY 5041 ACATAGTGGCCACTTCTGTCCAGACTAAACAGGACGAGAGCAACTTTCCTTACTCTGG 5100
Db 5041 ATATAGATGCGCCACTTCTATCCAGACAAAGCAGAGTGGGGAGACTTTCCTTACTCTGG 5100
QY 5101 TGGCATATCAAGCTTACAGTGTGGCGCCAGGCTCAAGTCCACCTCCATCTGTGGGACCAA 5160
Db 5101 TAGGTTACCAAGCAACGCTGTGGCTAGGGCTCAAGCCCTCCCGCATCTGTGGGACCAA 5160
QY 5161 TGTGGAGTGTCTCATAGCTGAGTGTGGCGCCAGGCTCAAGTCCACCTCCATCTGTGGGACCAA 5220
Db 5161 TGTGGAGTGTCTCATAGCTGAGTGTGGCGCCAGGCTCAAGTCCACCTCCATCTGTGGGACCAA 5220
QY 5221 GGCTAGGAGCGGCTCCAAATGAGGTCATCTCTACACACCCCAATCAATATCATATGG 5280
Db 5221 GACTGGGCGGTGTTTCAAGTGAAGTCACTGAGCGACCCCAATCAACCAATATCATATGA 5280
QY 5281 CATGCATGTGGGCTGAGCTGGAGGTCTGTACTAGCACTGGGCTCTGTAGGGGAGTCC 5340
Db 5281 CATGCATGTGGGCGGACCTGGAGGTCTGTACGAGCACTGGGTCCTGTGGGCGGCTCC 5340

QY	5341	TTGACGTTTGGCGGCATACTGCTCAGCACAGCAGTGTGGTCAATTGTGGGAGGATCA	5400
Db	5341	TGCGCTCTCTGGCGCGTATTGCTCTGTCAACAGCTGCGTGGTCA TAGTGGGAGGATTG	5400
QY	5401	TCCTGTCCGGGAAGCAGCTGTGCTTCCGACAGGGAATCTCTACAGGAGTTCCGATG	5460
Db	5401	TCCTGTCCGGGAAGCAGGCAATTATACCTGACAGGAGGTTCTCTACAGGAGTTCCGATG	5460
QY	5461	AGATGAAGAGTGTGCTCTCAACATTCCTTACATCGAGCAGGGAATGACGTCCGCGAGC	5520
Db	5461	AGATGAAGAGTGTCTCAGCATTTACCGTACATCGAGCAGGGAATGATGCTCGCTGAGC	5520
QY	5521	AATTCAAGCAAAAAGCGCTCGGTTTGTCAAAACGGCCACCAAGCAGGAGGAGTTCGTG	5580
Db	5521	AGTTCAAGCAGAAAGCGCTCGGCTCTCTGACGACCGCGTCCCGCCCAAGCAGAGTTATCA	5580
QY	5581	CTCCGCTGTGGAGTCCAAAGTGGAGGCCCTTGAGACCTTCTGGCGGACGACATGTGA	5640
Db	5581	CCCGCTGTCTCCAGACCAACTGCGAGAACTCTCAGGTCTTCTGGCGGAGACATGTGGA	5640
QY	5641	AATTTCATCAGCGGAATACAGTACTAGCAGGCTTATCCACTCTGCTCTGGAACCCCGCGA	5700
Db	5641	AATTTCATCAGTGGATACANATCTTGGCGGCGCTGTCAAGCTGCCTGGTAAACCCGCGA	5700
QY	5701	TAGCATCATTTGATGGCATTTACAGCTTCTATCACTAGCCGCTCAACACCCAAACACCC	5760
Db	5701	TTGCTTCAATTGATGGCTTTTACAGCTGCCGTCAACAGCCCACTAACCACTGGCCAAACCC	5760
QY	5761	TCCTGTTTAACTCTTGGGGGATGGTGCTGCCAACTCGCTCTCTCCAGCGCTGCGT	5820
Db	5761	TCCTCTTCAACATTTGGGGGGTGGGTGGCTGCCAGCTGCGCGCCCCCGTGCCTGTA	5820
QY	5821	CAAGTTTTCGTGGCGCGGCATCGCCGAGCGGCTGTGTGCAGCATAGGCTTTGGGAAGG	5880
Db	5821	CCGCTTTGTGGCGCTGGCTTAGCTGGCGCCGCCATCGGCAGCGTTTGGACTCGGGAAGG	5880
QY	5881	TGTCGTGGACATCTTGGCGGCTATGGGGCAGGGTAGCGGAGGCTACTGTTGGCCTTTTA	5940
Db	5881	TCCTCTGTGGACATCTTTCAGAGGTATGCGCGGCGCTGGCGGAGGCTCTTTGTAGCATCA	5940
QY	5941	AGGTCATAGCGCGGAGGTGCCCTCCACGAGGACCTGGTCACTTACTCCCTGCCATCC	6000
Db	5941	AGATCATAGCGGTAGGTCCCTCCACGAGAGACCTGGTCACTCTCTGCCCGCCATCC	6000
QY	6001	TCCTCTCTGTGTCCTGGTCTGTGGGTGCTGTGCGCAGCAATACCTGCTGCGCACTGG	6060
Db	6001	TCCTGCGCTGAGCGCTTGTAGTCGTTGTGCTGTGCGCAGCAATACTGCGCGCGCACGTTG	6060
QY	6061	GCCCGGAGAGGGGCTCTGCAGTGGATGACCGGCTGATAGCTTTCGCTTCGGGGGTA	6120
Db	6061	GCCCGGCGAGGGGCGAGTGATGGATGAACCGGCTTAATAGCCTTCGCTTCGGGGGGA	6120
QY	6121	ACCACGTCCTCCCTTACACACTATGTGCTCTGAGCGACGCTGTGACGACGCTGTCACTCAGA	6180
Db	6121	ACCATGTTTCCCCCAACACTACTGTCGCGGAGAGCGATGACGCGCCCGCTCACTGCCA	6180
QY	6181	TCCTCTCTAGCCTTACATCACTCAA CTGCTGAAGCGGCTCCACGATGGAATTAACGAGG	6240
Db	6181	TACTCAGCAGCCTCACTGTAAACCACTCTCTGAGCGCACTGCATCAGTGAATAGCTTCGG	6240
QY	6241	ACTGCTCTAGCCCATGCTCCGGCTCGTGGCTTAAGGGATGTTTGGATGGATATGCACGG	6300
Db	6241	AGTGTACACTTCCATGCTCCGGTTCCTGGCTTAAGGGACA CTGGGATCTGGATATGCGAGG	6300
QY	6301	TGTTGACTGACTTCAAGACCTTGGCTCCAGTCCAACTCCTGCCCGGTTACCGGGAGTCC	6360
Db	6301	TGCTGAGCGACTTTAAGACCTTGGCTGAAGCCACTCATGCCCACTGCTCTGGGATTC	6360
QY	6361	CTTTCCTGTCTATGCGCAACGCGGTTACAGGAGTCTTGGCGGGGGGACGGCATATGCAAA	6420
Db	6361	CTTTCGTGTCTCTGCGACGCGGTTATAGGGGGGCTCTGCGAGAGACGGCATATGACACA	6420

QY	6421	CCACCTGCCCATTCGGCAGACACAGATCCCGCGGACATGTCAAAAACGGTTCCATGAGGATCG	6481
DB	6421	CTCGCTGCCACTGTGAGGTGAGATCACTGGACATGTCAAAAACGGGACGATGAGGATCG	6481
QY	6481	TAGGGCTAGAACCTTCAGCAACACAGCTGCACGGAACGTTCCCATCAACGCAATACACCA	6541
DB	6481	TCGGTCTTAGGACCTGCAGGACATGTGGAGTGGGACGTTCCCATTTAACGCTTACACCA	6541
QY	6541	CGGGAAGCTTGACACACCTCCCGCGCCCCAACATTTCCAGGGCGCTATGCGGGTGCGTGG	6601
DB	6541	CGGGCCCCGTACTCCCTTCCTGCGCGCAACTAAGTTTCGCGCTGTGGAGGTGTCTGG	6601
QY	6601	CTGAGGAGTACGTGGAGGTTACCGCTGTGGGGGATTTCCACTAGTGACACGGGATGACCA	6661
DB	6601	CAGAGGAATACGTGGGAGTAAGGCGGGTGGGGACTTCCACTACGTATCGGGTATGACTA	6661
QY	6661	CTGACAACTGAAAGTGCACATGCCAGGTTCCGGCCCCCGAAATTTCTCAAGAGAGTGGATG	6721
DB	6661	CTGACAACTTTAAATGCCGTGCCAGATCCCATCGCCCGAAATTTTTCACAGAAATTGACG	6721
QY	6721	GAGTGGGGTTGACACAGTACGCTCCGGCGGTGCAAACTCTTCTTACGGGAGGAGTCACGT	6781
DB	6721	GGGTGCGCTTACATAGGTTTGGCGCCCCCTTGCAGCCCTTGCTCGGGAGGAGTATCAT	6781
QY	6781	TCAGGTGCGGCTCAACCAATATCTTGGTCCGCTCGAGTCCCATGGGACCCGACCCGG	6841
DB	6781	TCAGAGTAGGACTCCACAGTACCGCGTGGGGTCGCAATTAACCTTTCGAGCCCGCAACCGG	6841
QY	6841	ACGTAAAGTGCTTACTTCCATGCTCACCGATCCCTCCCACTTACAGCAGAGACGACTA	6901
DB	6841	ACGTAGCCGTGTTGAGCTCATGCTCACTCATCTCCCTCCCATTAACAGCAGAGGCGCGG	6901
QY	6901	AGCCTAGGCTGGCTAGAGGGTCTCCCCCTCTTTTAGCCAGCTCATCAGCTAGCCAGTTGT	6961
DB	6901	GGAGAGGTTGGCAGAGGGTCAACCCCTCTATGCGCCAGTCTCCTCGCCAGCCAGCTGT	6961
QY	6961	CTGCGCTTTTGAAGCGACATGCACTACCAACCATGACTCTCCCGCGAAGCTCGTCACTCA	7021
DB	6961	CCGCTCATCTCTCAAGGCAACTGACCGCCCAACCATGACTCTCCCTGAGCCCGAGCTCA	7021
QY	7021	TCGAGGCCAACCTTTGTGCGCGCAGAGATGGCGGAACATCACCTCGCGTGGAGTCAG	7081
DB	7021	TAGAGGTAATACCTCTGTGAGGACAGAGATGGCGGCACATCACAGGGTTGAGTCAG	7081
QY	7081	AGAATAAGGTAGTAATTTCTGCACTCTTCGAAACCGCTTCACGGCGAGGGGATGAGAGGG	7141
DB	7081	AGAACAAAGTGGTGAATTTCTGCACTCTTCGACTCCCTTTGTGGCAGAGGATGAGCGGG	7141
QY	7141	AGATATCCGTGGCGGCGGAGATCTCGGAAATTCAGGAAGTTCCCTCTAGCGTTGCGCCA	7201
DB	7141	AGGTCTCCGTACCCGCAAGAAATTTCTGGGAAGTCTCGGAGATTCGCGCGGCGCTTCGCG	7201
QY	7201	TATGGGACCGCCGGACTACAACTCTCCACTGCTAGAGTCTCTGGAAGACCCCGACTACG	7261
DB	7201	TTTGGGCGCGCGGACTACAAACCCCGCTGTAGAGAGCTGGAAGAACCTGACTACG	7261
QY	7261	TCCCTCCGCTGGTACACGGATGCCATTCGCCACTACGAAGGTCCTTCCCAATACCACTTC	7321
DB	7261	AACCACTGTGGTCCATGGCTGACCGGTACCACTCTACGGTCCCTCTGTGCTTCGCTCGG	7321
QY	7321	CACGGAAGAGAGACGGTTGTCTGTGACAGAAATCCAGAAATTCCTGTCTTCTGCGTGGCGAGC	7381
DB	7321	CTCGAAAAAGCGTACGGTGGTCTCTACCGCAATCAACCTTATCTACTGCTTGGCCGAGC	7381
QY	7381	TCGCCACTAAGACCTTCGGTAGCTCCCGATCGTGGCGGTTGTATGACGGGACCGCGGACCG	7441
DB	7381	TTGCCACCAAAAGTTTGGCAGCTCTCTCACTTCGGCATTTACGGGCGCAATACGACAA	7441
QY	7441	CCCTTCTGACCTCGGCTCCGACGAGCGGTGACAAAGGATCCGAGCTTTCAGTCGCTACTCCT	7501
DB	7441	CATCCTCTGAGCCCGCCCTCTCTGGCTGCCCCCCCCGACTCCGAGCTTGAGTCTTATCTT	7501
QY	7501	CCATGCCCCCCCTTTGAAGGGGAGCCGGGGACCCCGATCTCAGCGACGGGCTTTGGTCTA	7561

Db 7501 CCATGCCCCCTCGAGGGGAGCCTGGGATCCGGATCTCAGCGACGGGTATGTCGA 7560
 QY 7561 CCGT---GAGTGGAGGCTGAGGATGTCGTCTGCTGCTCAATGTCCTATACCTGA 7617
 Db 7561 CCGTCACTAGTGGGGCCGACACGGAAGATGTCGTCTGCTCAATGTCCTATACCTGA 7620
 QY 7618 CAGGCGCCCTGATCAGCCATGGCTGCGAGGAAGTAAAGCTGCCCATCAACCCCTTGA 7677
 Db 7621 CAGGCGACTGCTCAACCCGTGGCTGCGGAAGAACAAAACTGCCCCATCAACGCACTGA 7680
 QY 7678 GCACACTTTTGTCTGCTCAACCAACATGCTACGCCCAACATCCCGAGCGGAAGCC 7737
 Db 7681 GCAACTCGTTGCTACGCCATCAAACTGCTGATTCACCACTTCACGCACTGCTGCC 7740
 QY 7738 TCCGCGCAGAGAGGTCACCTTTGACAGATGGAAGTCTGGATGATCATTCACGGGAGC 7797
 Db 7741 AAGGCGAAGAAAGTCAATTTGACAGACTGCAAGTCTGGACAGCCATCACGAGAGC 7800
 QY 7798 TACTCAAGAGATCAAGCGAAGCGTCCACAGTTAAGGCTAAGCTTCTATCTATAGAGG 7857
 Db 7801 TGCTCAAGAGGTCARAGCAGCGGTCAAAAGTGAAGGCTAATCTGCTATCCGTAGAGG 7860
 QY 7858 AGGCTCGAAGCTGACGCCGCCACATTCGGGCAAAATCCAAATTTGGCTATGGGGCAAGG 7917
 Db 7861 AAGCTTCAGCCTGACGCCGCCACATTCAGCCAAATCCAAAGTTTGGCTATGGGGCAAAAG 7920
 QY 7918 ACCTCCCGAACCTATCCAGAGGCGCTTAACACATCCGCTCGTGTGGAGGACTTC 7977
 Db 7921 ACCTCCCGTTGCTATGCCAGAAAGCGCTAGCCCACTCACTCCGTTGGAAGACCTTC 7980
 QY 7978 TGGAGACACTGAACACCAATTAACACACCATCATGGCAAAAGTGAAGTTTCTGCG 8037
 Db 7981 TGGAGACAGTGAACCAATTAACACTACATCATGGCCAAAGACAGAGTTCTGCG 8040
 QY 8038 TCCAAACAGAGAGGAGCGGCAAGCAGCTCGCTTATCGTATTCAGACCTGGAG 8097
 Db 8041 TTCAGCTGAAGGGGGTCTGAAGCAGACTCTCTCATCTGTTCCCGACCTGGCG 8100
 QY 8098 TTCTGCTATGCGAAGATGCGCTTTACAGAGTGGTCTCCACCTTCCTCAGCGCGTGA 8157
 Db 8101 TGCCTGCTGCGAAGATGCGCTTTACAGAGTGGTCTAGCAAGTCCCTCGCGCTGA 8160
 QY 8158 TGGCTCTCTCATACGATTTCAATATCTCCCAAGCAGCGGTGAGTCTGCTGGAATA 8217
 Db 8161 TGGGAAGCTCTACGGATTCATATCTCACCAGACAGCGGTTGAATTCCTCGTGAAG 8220
 QY 8218 CCTGGAATCAAGAAATGCGCTATGGCTTCTATATGACACCGCTGTTTGACTCAA 8277
 Db 8221 CGTGGAGTCCAAAGAACCCCGATGGGTTCTGATGATACCGCTGTTTGACTCCA 8280
 QY 8278 CGTCACTGAGAGTGACATTCGTGTGAGGAGTCAATTTACCAATGTTGACTTGGCCC 8337
 Db 8281 CAGTCACTGAGAGCGACATCCGTACGAGAGGAGCAATTTACCAATGTTGACTGGACC 8340
 QY 8338 CCGAGCCAGACAGGCCATAGGTCGTCAAGAGCGGCTTCAATCGGGGTCCCTGA 8397
 Db 8341 CCCAAGCCCGGTGCCATCAAGTCCCTCACTGAGAGGCTTATGTTGGGGCCCTCTTA 8400
 QY 8398 CTAACTCAAGAGGAGAGCTGCGTTATTCGCCGTGCGCGCAGTGGCGTGTGAGGA 8457
 Db 8401 CCAATTCAGAGGGGAAACTGCGGCTACCGCAGTGGCGGCGAGCGGTACTGACAA 8460
 QY 8458 CTAGCTGGGTAATACCTTCATGTTACTTGAAGGCCACTCAGCCCTGTGAGCTGCAA 8517
 Db 8461 CTAGCTGTGGTAAACCCCTCACTTGTATCATCAAGCCCGGCGAGCTGTGAGCCGAG 8520
 QY 8518 AGCTCAGAGCTGACAGATGCTGCTGAACGAGAGACAGCTTGTCTGTTATCTGTGAAGCG 8577
 Db 8521 GGCTCCAGGACTGCACCAATGCTGCTGTGTTGGGACGAGCTTAGTCTGTGAAGATG 8580
 QY 8578 CCGGAACCCAGAGAGTGGGGGCGCCCTACGAGCCTTACAGGAGCTATGACTAGTATT 8637

Db 8581 CCGGGCTCAGAGGACGCGGGGAGCCTTGAGAGCCTTCAOGGAGGCTATGACCAAGTACT 8640
 QY 8638 CCGCCCCCOCGGGATCCGCCCAACAGAAATACGACTCTGGAGCTGATAACATCATGTT 8697
 Db 8641 CCGCCCCCOCGGGAGCCGCCCAACAGAAATACGACTCTGGAGCTTATAACATCATGCT 8700
 QY 8698 CTTCAATGTTGTCAGTGGGACAGATGATCTGTCGCAAAAGGATGATACCTCACCCGTG 8757
 Db 8701 CTTCAACAGTGTGTCGCGCCACGACGCGCTGGAAGAGGGGTCTACTACCTTACCCGTG 8760
 QY 8758 ACCCAACACCCCTTGCACGGGCTGGTGGGAGACAGCTAGACACACTCCAACTCACT 8817
 Db 8761 ACCCTACACCCCTTGCAGAGAGCGCTGGGAGACAGCAAGACACACTCCAGTCAAT 8820
 QY 8818 CTTGGCTAGGCAATATCATGATGATGTCGTCGTCACCTATGGSCAAGGATGATCTGATGA 8877
 Db 8821 CTTGGCTAGGCAATATCATGATGATGTCGTCGTCACCTATGGSCAAGGATGATGATGA 8880
 QY 8878 CTCACCTTTTCTCCATCTCTAGCTCAAGAGCAACTTGAAGGCTTGGATTTGTCAGA 8937
 Db 8881 CCATTTCTTTAGCGTCTCATAGCAGGATCAGCTTGAAACAGGCTCTTAATCTGTAGA 8940
 QY 8938 TCTACGGGCTTGTACTCCATTTGAGCCACTTGAOCTACCTCAGATCATTTGAACAGCTCC 8997
 Db 8941 TCTACGGGCTTGTACTCCATTTGAGCCACTTGAOCTACCTCAGATCATTTGAAGACTCC 9000
 QY 8998 ATGGTCTTAGGCAATTTAATCTCAGAGTTACTCTCCAGTGAGATCAATAGGTTGCTT 9057
 Db 9001 ATGGCTCAGCGCATTTTCACTCCACAGTTACTCTCCAGTGAATCAATAGGTTGGCG 9060
 QY 9058 CATGCTCAGGAACTTTGGGGTACACCTTGGACCTTGGAGACATCGGGCCAGAGTG 9117
 Db 9061 CATGCTCAGGAACTTTGGGGTACACCTTGGAGCTTGGAGACATCGGGCCAGAGCG 9120
 QY 9118 TCCGCTAAGCTTACTGTCOCAGGGGGAGGCGCCACTTGTGTCAGATCACTCTTGA 9177
 Db 9121 TCCGCTAAGCTTACTGTCOCAGGGAGGCGAGGCTGCCATATGTGGCAAGTACTCTTCA 9180
 QY 9178 ACTGGGAGTAAAGCAAGCTTAAACTCACTCAATCCGCGCGCTCCAGCTGAGCT 9237
 Db 9181 ACTGGGAGTAAAGCAAGCTTAAACTCACTCAATCCGCGCGCTCCAGCTGAGCT 9240
 QY 9238 TGTCTGCTGTTGCTGCTGCTTACAGCGGGGAGACATATATCAGAGCTCTCTGCTG 9297
 Db 9241 TGTCTGCTGTTGCTGCTGCTTACAGCGGGGAGACATATATCAGAGCTCTCTGCTG 9300
 QY 9298 CCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9357
 Db 9301 CCGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9360
 QY 9358 TGTCTCCCAACCGATGAACGGGAGCTTAACCACTCCAGGC-CITTAAGCCATTTCTGTTT 9416
 Db 9361 TCTTCCCAACCGATGAAGGTTGGGTAACACTCCGCGCTCTTAGGCCATTTCTGTTT 9420
 QY 9417 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9476
 Db 9421 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9480
 QY 9477 TT-----TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9489
 Db 9481 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9540
 QY 9490 TCTTTAATGTTGCTCCATCTTAGCCCTAGTCAAGGCTAGCTGGAAGGTCCTGAGAGC 9549
 Db 9541 TCTTTAATGTTGCTCCATCTTAGCCCTAGTCAAGGCTAGCTGGAAGGTCCTGAGAGC 9600
 QY 9550 GCATGACTGACAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9595
 Db 9601 GCAATGACTGACAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9646


```

; Sequence 1, Application US/09995937
; Publication No. US2003002010A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
;
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995 937
; FILING DATE: 28-Nov-2003
; PRIORITY INFORMATION:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-995-937-1

```

```

Query Match      65.2%; Score 6254.8; DB 10; Length 9646;
Best Local Similarity 78.4%; Pred. No. 0; Mismatches 2037; Indels 51; Gaps 3;
Matches 7558; Conservative 0;

1  GCCAGCCCCCTGATGGGGGCGACACTCCACCAATGAATCACTCCCTGTGAGGAATCTACTG 60
1  GCCAGCCCCCTGATGGGGGCGACACTCCACCAATGAATCACTCCCTGTGAGGAATCTACTG 60
61  TCTTTCAGCAGAAAGCGCTACCCATGCGGTAGTATGATGATGATGATGATGATGATGATGAT 120
61  TCTTTCAGCAGAAAGCGCTACCCATGCGGTAGTATGATGATGATGATGATGATGATGATGAT 120
121  CCCCCCTCCGGAGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGGAATTCGCAG 180
121  CCCCCCTCCGGAGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGGAATTCGCAG 180
181  GACGACCGGGTCTCTTCTTGATCAACCGCTCAATGCTCGGAGATTTGGGGGTGCCCC 240
181  GACGACCGGGTCTCTTCTTGATCAACCGCTCAATGCTCGGAGATTTGGGGGTGCCCC 240
241  GCGAGTCTCTAGCCGAGTGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 300
241  GCAAGACTCTAGCCGAGTGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 300
301  GTGCTTGGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 360
301  GTGCTTGGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 360

```

```

361  CTCAAGAAAAAACAACCTAACAACAACCGCGGCCACAGAGAGCTCAAGTTCCCGGGCG 420
361  CTCAAGAAAAAACAACCTAACAACAACCGCGGCCACAGAGAGCTCAAGTTCCCGGGCG 420
421  GTGGTCAGATCGTTGGTGGAGTTTACTGTCGCGGCGAGGGGCCCGAGTTGGGTGTGC 480
421  GCGGTGAGATCGTTGGTGGAGTTTACTGTTGTCGCGGCGAGGGGCCCGAGTTGGGTGTGC 480
481  GCGCGACTAGGAAGGCTTCGAGCGGTGCAACCTCTGTGGAAGGCGACAACTATCCCAA 540
481  GCGCGAGAGAGAGACTTCGAGCGGTGCAACCTCTGTGGAAGGCGACAACTATCCCAA 540
541  AGCTTCGCGACCCGAGGAGGAGGCTTCGAGCGGTGCAACCTCTGTGGAAGGCGACAACTATCCCAA 600
541  AGGCACTGCGCGCGGAGGAGGAGGCTTCGAGCGGTGCAACCTCTGTGGAAGGCGACAACTATCCCAA 600
601  GCAATGAGGCTTGGGGTGGCGAGGAGGCTTCGAGCGGTGCAACCTCTGTGGAAGGCGACAACTATCCCAA 660
601  GCAATGAGGCTTGGGGTGGCGGAGGAGGCTTCGAGCGGTGCAACCTCTGTGGAAGGCGACAACTATCCCAA 660
661  GGGGCCCCACGAGACCCCGCGGTAGGTGCGGTAACTTGGTAAAGTCAATGATCCCTTAA 720
661  GGGGCCCCACGAGACCCCGCGGTAGGTGCGGTAACTTGGTAAAGTCAATGATCCCTTAA 720
721  CATCGCGCTTCGCGGATCTCATGGGTACATTCGGCTCGTGGGCGCCCGCTAGGGGGCG 780
721  CGTGGGCTTCGCGGATCTCATGGGTACATTCGGCTCGTGGGCGCCCGCTAGGGGGCG 780
781  CTGCGAGGCTTCGCGGATCTCATGGGTACATTCGGCTCGTGGGCGCCCGCTAGGGGGCG 840
781  CTGCGAGGCTTCGCGGATCTCATGGGTACATTCGGCTCGTGGGCGCCCGCTAGGGGGCG 840
841  GGAACCTTCGCGGATCTCATGGGTACATTCGGCTCGTGGGCGCCCGCTAGGGGGCG 900
841  GGAACCTTCGCGGATCTCATGGGTACATTCGGCTCGTGGGCGCCCGCTAGGGGGCG 900
901  TCCGAGCTTCGCGGATCTCATGGGTACATTCGGGTACATTCGGGTACATTCGGGTACATTCGGGT 960
901  TCCGAGCTTCGCGGATCTCATGGGTACATTCGGGTACATTCGGGTACATTCGGGTACATTCGGGT 960
961  GCTCCAACTCAAGCATTTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
961  GCTCCAACTCAAGCATTTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
1021  TGCCCTGTGTTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
1021  TGCCCTGTGTTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
1081  GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
1081  GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
1141  GGAGGCTGTCTTCTGCTCCGCTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
1141  GGAGGCTGTCTTCTGCTCCGCTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
1201  TCTCCAGCTGTTCACTTCTGCTCCGCTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
1201  TCTCCAGCTGTTCACTTCTGCTCCGCTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
1261  CAATCTATCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
1261  CAATCTATCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
1321  CACCTTACAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
1321  CACCTTACAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
1381  TGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
1381  TGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
1441  ACTGGGCTAAGGTTCTGATTTGTTGGGCGTACTCTTTTGGGCGGCTTACCGGGGAGAGCCACA 1500

```

1441 ACTGGCGAAGGTCCTGTAAGTCTGCTAATTTGCCGCGTCGACGCGGMAACCCACG 1500
1501 CGACGGGGAGGGTGGCGGCGCAACACACCTCCGGGTTCAAGTCCCTTTCTCATCTGGGG 1560
1501 TCACCGGGGAAAGTGGCGGCGGACACACGCTGGGCTTGTGGTCTCTTTACACCAAGCG 1560
1561 CGTCTCAGAAATCCAGCTTGTGAATACCAACGCGACGCTGSCACATCAACAGAGACTGCC 1620
1561 CCAAGCAGAAATCCAACTGATACCAACGCGAGTTGGCACAATCAATAGACAGGCGCT 1620
1621 TAAATTCGAATGACTCCCTCAAACTGGGTCTTTGGCGGCGTGTTTAGCGACACAAGT 1680
1621 TGAATTCGAATGAAAGCTTTAAACACCGCTGGTTAGCAGGCTCTTCTATCAGCACAAT 1680
1681 TCAACTCGTCGGGTGGCGGAGCGCATGGCAGCTGCGGCCCAATGACTGGTTGCC 1740
1681 TCAACTCTTCAGGCTGTCCTGAGAGGTGGCGCAGCTGCCAGCGCTTACCGATTTTGCC 1740
1741 AGGGTGGGCCCCATCACTATACCTAAGCTTAACGCTCGGATCAGAGCGCTTATTGCT 1800
1741 AGGCTGGGTCCTATCAGTTATGCCAAAGGAGCGCTCGACGAAGGCCCTACTGCT 1800
1801 GGCAATACGGCTCGACCGCTGCTGCTGACCGGCTGCGAGGTGTGCTCCAGTGT 1860
1801 GGCACTACCTCCCAAGACCTTGTGGCATTTGTCCGCAAGAGCGTGTGTGGCCGCTAT 1860
1861 ATTGTTTACCCCAAGCGCTTGTGTGGGACACGATCTGTCGGTGTCCCTACGT 1920
1861 ATTGCTTCACTCCAGCGCCCTGTGTGGGAGCAGCAGCTCGGCGCGGCTACCT 1920
1921 ATAGCTGGGGAGAAATGAGACAGCTGATGCTCTCAACAAACAGCGCTCCGCCACAAG 1980
1921 ACAGCTGGGTGCAATGATGATGATGCTCTGCTTCCCTTAAACACAGCGCCACCGCTG 1980
1981 GCAACTGGTTCGGGTGACATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
1981 GCAATGGTTCGGGTGACATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
2041 CCCCGTGAACATCGGGGGGTGCGTAACCGACCTGATGATGATGATGATGATGATGATG 2100
2041 CCCCTTGTGTCATCGGAGGGGTGGGCAACACCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
2101 GGAAGCACCACCGCTACTTACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
2101 GCAAGCATCCGGAAGCCACATCTCTCGGTGGGTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
2161 GCCTAGTAGACTACCCATACAGGCTTTGGCAGTACCCCTGCACTCTCAATTTTCCATCT 2220
2161 GCATGGTGCATACCCGCTATAGGCTTTGGCAGTATCTTGTGACCAATACACCATAT 2220
2221 TTAAGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280
2221 TCAAGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280
2281 CTCGAGGAGAGCGGTGTAATTTGGAGGACAGGATGATGATGATGATGATGATGATGATG 2340
2281 CGCGGGCGAAGCGTGTGATCTGGAGAGACAGGACAGGCTCGAGCTCAGCCCATTTGCTG 2340
2341 TGTCTAACAGAGTGGGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
2341 TGTCCACACACAGTGGGAGGCTCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
2401 CTGGTTTGCATCTCCATCAGAACATCTGGAGCTGCAATACCTGATGAGGCTGAGGCT 2460
2401 CCGGCTCATCCACCTCCACAGAACATCTGGAGCTGCAATGATGATGATGATGATGATG 2460
2461 CAGCGTTTGTCTCTTTGCAATCAATGGAGTACATCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
2461 CAAGCACTCGGCTCTGGGCAATTAAGTGGGAGTACGCTGCTGCTGCTGCTGCTGCTGCT 2520
2521 CAGACGGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580

2521 CAGACGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580
2581 CTTTAGAGAACTTGTGCTGCTCAATGCGGCTGCGTGCAGGAGCGATGATGATGATGCT 2640
2581 CTTTGGAGAACTTGTGATTAATCAATGAGCAATCTGCGCGGAGCGACAGGCTGCTGCT 2640
2641 CTTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
2641 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
2701 CGTATGCTTTTATGCGGATGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760
2701 TCTACGCTTCTACGGAATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760
2761 CTTACGCTTGGACCGGAGATGCGTGCATGCTGCGGGGTGCGTCTGCTGCTGCTGCTGCT 2820
2761 CATACGCACTGGACCGGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820
2821 TATTTCTGACCTTGTCAACATACATAAAGTGTCTCTCTAGGCTCATATGCTGCTTAC 2880
2821 TGGCGCTGACTCTGCTGCGCATATTAAGCGCTACATCAGCTGGTGCATGCTGCTGCTGCT 2880
2881 AATACTTTATACACAGACCGGAGCGCATGCAAGTGTGGGTGCTGCTGCTGCTGCTGCTGCT 2940
2881 AATTTTCTGACGAGTAGAAGCGCACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940
2941 GGGAGGCGCGGATGCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
2941 GGGGGGGGCGGATGCGCTCACTTACTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
3001 ACATACCAAACTCTGCTGCGCATCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3060
3001 ACATACCAAACTCTGCTGCGCATCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3060
3061 CGAGATGCGGCTGCTGCTGCGGCTCAAGGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCT 3120
3061 TTAAGTTCCTGCTGCTGCGGCTCAAGGCTCTTCCGATCTGCGGCTGCTGCTGCTGCTGCT 3120
3121 AAGTTCGCGGGGTCAATATGCTCAAAATGCTCTTCAAGCTGGGCGGCTGCTGCTGCTGCT 3180
3121 AGATGCGGAGGCTCATGCTGCAATGCTGCAATGCTGCAATGCTGCAATGCTGCTGCTGCT 3180
3181 CGTACGTTTATACCACTTTTACCCCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCTG 3240
3181 CCTATGCTATACCACTCTCACCCTCTTCCGACCTGCTGCGGCTGCTGCGGCTGCTGCTG 3240
3241 TTCGCTGCGGCTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3300
3241 TGGCGGTGGCTGAGAACCTGCTGCTTCTTCCGATGAGAGACCAAGCTCATCAGCTGCTG 3300
3301 GAGCAGACACCGCTGCTGCTGCGGACATCATCTTGGGCTTACCCGCTGCTGCTGCTGCTG 3360
3301 GGGCAGATACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3360
3361 GGAAGGAGATATTTTGGGACCGGCTGATGCTGCTGAGGCGAGGCTGCTGCTGCTGCTGCT 3420
3361 GCCAGGAGATGCTGCTTGGGCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3420
3421 GGGCCTATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3480
3421 GGGCCTATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3480
3481 TCACAGGCGGCGGACAGAACAGCTGCTGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3540
3481 TGAATGCGCGGACAAACCAAGTGGAGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3540
3541 AATCTTCTGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
3541 AATCTTCTGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
3601 CGAGAGACCTTAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3660
3601 CGAGAGACCTTAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3660

5821 CCGCCCTTGTGGCGCTGGCTTACGTGGCGCCCATCGACGCTGGACTGGGGAAG 5880
5881 TGTCTGTGACATCTTGGCGGGCTATGGGCGACGGGTGACCGCGCACTCGTGGCCCTTAA 5940
5881 TCCTCTGGGACATCTTGCAGGGATATGGCGCGGGCTGGCGGAGCTCTGTAGCAATCA 5940
5941 AGGTCAATGACGGCGAGGTCCCTCCACCCAGAGACCTGCTCAACTTACTCCCTGCCATCC 6000
5941 AGATCATGACGGTGTAGTCCCTCCACCGAGGACCTGCTCAACTTACTCCCTGCCATCC 6000
6001 TCTCTCTGTGTCCTGTGGTCTGGGGTGTGTGCGGAGCAAACTACTGCGTGGGACGTGG 6060
6001 TCTCGCTGGAGCCCTTGTAGTCGGTGTGTGCTGCGCAGCAATACTGCGCGCGCACTTG 6060
6061 GCCCGGAGAGGGGCTGTCCAGTGTGATGAACCGGCTGTAGAGCTTCCGCTTCCGGGGTA 6120
6061 GCCCGGCGAGGGGGAGTGAATGGAATGAACCGGCTTAATAGCTTCCGCTTCCGGGGGA 6120
6121 ACCACGTCTCCCTVACGCACTATGTGCTTGAGAGCGACGCTGCGAGCACTGTCACTCAGA 6180
6121 ACCATGTTTCCCCACGCACTAAGTGGCGGAGCGATGCGCGCGCGCTCACTGCCA 6180
6181 TCTCTCTAGCTTACCATCTCAACTGCTGAGAGCGGCTCCACGAGTGTATGAGG 6240
6181 TACTCAGACGCTCACTGTAAACCAAGCTCTGAGGCGACTGCATCACTGGATAGCTCGG 6240
6241 ACTGCTCTAGCCATGTCCGGCTCGTGGCTTAAGGGATGTTGGGATTTGGATATGACCG 6300
6241 AGTGTACCATCTCAATGCTCCGGTCTCTGGCTTAAGGACATCTGGGACTGTGATGCGAG 6300
6301 TGTGACTGACTTCAAGACCTGGCTCCAGTCCAACTCTGCGCGGTTAACCAGGATCC 6360
6301 TGCTGAGCGACTTAAAGACCTGGCTGAAGCAAGCTCATGCCACAAGCTGCTGGGATC 6360
6361 CTCTTCTGTCTATGCAACCGGGTACAAAGGAGTCTGGCGGGGAGCGGATCATGCAAA 6420
6361 CCTTGTGTCTGTCAGCGCGGGTATAGGGGGTCTGGCGGAGAGCGGCATTTAGCACA 6420
6421 CCACCTGCCATCGGAGCAGATCGCGGACATGTCMAAAACGTTCCATGAGGATCG 6480
6421 CTCGCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAAAACGAGCAATGAGATCG 6480
6481 TAGGGCTTAGAACCTGCGAGCAACAGTGGCAACGAACTTCCCATCAACGCATACACCA 6540
6481 TCGGTCTTAGACCTGACGAGCAATGTGGAGTGGACGCTTCCCATTAACGCTACACCA 6540
6541 CCGGACCTTGACACCTCCCGGGGCGCAACTATTCAGGGCGCTATGGCGGTGGCTG 6600
6541 CCGGCGCTGTACTCCCTTCTGCGCGCACTATAAGTTGCGGCTGTGGAGGGTGTCTG 6600
6601 CTGAGGAGTACGTGGAGGTTACGGCTGTGGGGGATTTCCACTAGCTGACGGGCTAGCCA 6660
6601 CAGAGGAATACTGGAGATAGCGGGTGGGGACTTCCACTAGTATCCGGTATGACTA 6660
6661 CTGACAAAGTAAAGTCCCATGCGAGTTCGGGCGCCCGAAATCTTCCAGGAGTGGATG 6720
6661 CTGACAACTTAAATGCGCGTCCAGATCCCATCGCCCGAAATTTTTCAGAAATTTGACG 6720
6721 GAGTCCGCTGACAGGTACCTCCGGGCTGCAAACTCTTCTAGCGGAGGACGTACGT 6780
6721 GGGTCCGCTCATATAGGTTTGGCGCCCTTTCAGAGCCCTTGTGCGGAGGAGGTATCAT 6780
6781 TCCAGGTCGGGCTCAACCAATCTTGTGGGTGCGAGCTCCCATGCGAGGCCGCAACCGG 6840
6781 TCAGAGTAGGACTCCACGAGTACCGGGTGGGGTGGCAATTAACCTTGGAGCGCCGAACCG 6840
6841 ACGTAAAGTGTACTTCCATGCTCAACGATCCCTCCCACTATACAGAGAGAGCGCTA 6900
6841 ACGTAGCGGTGTGAGTCCATGCTCACTGATCCCTCCCATATACAGAGAGGCGCGG 6900
6901 ACGTAGGCTGCTAGAGGCTCCCTCCCTCTTTAGCCAGCTCATCAGTACCCAGTGTGT 6960

6901 GGAGAAAGTGTGGGAGAGGGTCAACCCCTTCTATGCGCAGCTCTCCGCCAGCCAGCTGT 6960
6961 CTGCGCTCTTCTTGAAGGCGCATGCACTACCCACATGACTCCCGGAGCGCTGACCTCA 7020
6961 CCGCTCCATCTCTCAAGGCACTTTCACCGCCACCATGACTCCCTGACCGCGAGCTCA 7020
7021 TCGAGCCCAACCTCTTGTGGCGGAGGAGATGGGGGAAACATCACTCGCTGGAGTCA 7080
7021 TAGAGGCTAACCTCTGTGGAGGAGAGTGGCGGCAACATCAACAGGTTGAGTCA 7080
7081 AGAATAAGGTAGTAATCTTGGACTCTTTCGAAACGCTTCAACCGGAGGGGATGAGAGG 7140
7081 AGAACHAAGTGTGATCTTGGACTCTTTCGATCCGCTTGTGAGAGAGGATGAGCGGG 7140
7141 AGATATCCGTCGCGGCGGAGATCTTTCGAAATTCACAGGAAGTTCOCCTCAGGTTGCCCA 7200
7141 AGGTCTCCGTACCGGAGAAATCTTGGGAGTCTCGGAGATTCGCGCGGCTGCGCG 7200
7201 TATGGGACGCGCGGACTACATCTTCCACTGCTAGAGTCTTGGAGAGACCGGACTACG 7260
7201 TTTGGGCGCGCGGACTACACCCCGCTAGTAGAGACGTGGGAAAGACCTGACTACG 7260
7261 TCCCTCCGCTGTACACGATGCCATTCGCACTACCAAGGCTCTTCCAAATACCACTTC 7320
7261 AACCACTGTGTCTCATGCTGCCGCTACACCTCCACGCTCCCTCTGTCCTCCG 7320
7321 CACGAGAAAGGAGGACGTTGCTGACAGATTCATGTGCTTCTGCTTGGCGGAGC 7380
7321 CTGGAAAGAGTACGCTGGTCTTCCAGCAATCAACCTATCTACTGCTTGGCGAGC 7380
7381 TCGGCACTAAGACCTTCCGTTAGCTCGGATCGTGGCGGTTGATAGCGGACGCGCACCG 7440
7381 TTGCCACCAAAAGTTTGGGAGCTCTCACTTCCGCGATTTAGGGCGCAATACAGCAA 7440
7441 CCCTTCTGCTACCTCGGCTCCGACGAGTGCACAAAGGATCCGACGTTGAGTGTACTCCT 7500
7441 CATCTCTGAGCCGCCCTTCTGGCTGCCCCCGACTCCGACGTTGAGTGTACTTCT 7500
7501 CCATGCCCCCTTGAAGGGGAGCGCGGGGACCCCGATCTCAGGACGGGCTTGTGCTA 7560
7501 CCATGCCCCCTTGGAGGGGAGCTTGGGATCCGGATCTCAGCGAGGGTCTATGGTCA 7560
7561 CCGT---GAGTGGAGGCTAGTGGATGCTGCTGCTGCTCAATGCTCTATACCTGGA 7617
7561 CCGT---GAGTGGAGGCTAGTGGATGCTGCTGCTGCTCAATGCTCTATACCTGGA 7620
7618 CAGCGCCCTGTATCAACGCAATGCGCTCGGAGGAAAGTAAAGCTGCCCATCAACCCGTTGA 7677
7621 CAGGCGCACTCGTCAACCCGCTGCGGAGAAACAAACCTGCCATCAACGCACTGA 7680
7678 GCAACTCTTGTGCTGCTACCAACAATGCTGTAGGCAACAATCCCGCAGCGCAAGCC 7737
7681 GCAACTCTTGTGCTGCTACCAACAATGCTGTAGGCAACAATCCCGCAGCTGCTTGGC 7740
7738 TCGGCAAGAGAGGCTCACTTGTGACAGATTCGAAAGTCTGGATGATCATTCAGGAGC 7797
7741 AAGGCAAGAAAGTCACTTGTGACAGTGTGAGTCTTGGACAGCAATTAACGAGCG 7800
7798 TACTCAAGGAGTGAAGCGGAGCGCTCCACAGTTAAGGCTTAAGCTTCTATATAGAGG 7857
7801 TGCTCAAGGAGTCAAGAGCGCGCTCAAAAGTGAAGGCTAACTTGTCTATCCTAGAGG 7860
7858 AGGCTCGAAGTGAAGCGCGGCACTTCCGCGCAATTCGAAATTTGGCTATGGGCAAGG 7917
7861 AAGCTTCAGCGCTGACCGCGGCACTTCCAGCAATTCGAAATTTGGCTATGGGCAAGG 7920
7918 ACGTCCGGAACCTATCCAGCAGGCGCGTTAAACCACTCCGCTCCGTTGGGAGGACTTGC 7977
7921 ACGTCCGTTCCATGCGGAGAAAGCGCTAGCGCACTCACTCCGTTGGGAGGACCTTC 7980
7978 TGGAGACACTGGAACCAATTCACACCACTTCAATGGCAAAAGTGAAGTCTTCTGCG 8037
7981 TGGAGACAGTGTAAACCAATAGACACTTACCATCATGGCCAGAACGAGGTTTCTGCG 8040


```

SEQUENCE CHARACTERISTICS:
  LENGTH: 9646 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-917-563-1

Query Match      65.2%; Score 5254.8; DB 10; Length 9646;
Best Local Similarity 78.4%; Pred. No. 0;
Matches 7558; Conservative 0; Mismatches 2037; Indels 51; Gaps 3;

QY 1 GCCAGCCCCCTGATGGGGGACACTCCACCATGAATCACTCCCTCTGTGAGGAACACTCTG 60
DB 1 GCCAGCCCCCTGATGGGGGACACTCCACCATGAATCACTCCCTCTGTGAGGAACACTCTG 60

QY 61 TCTTACGACAGAAAGCGCTTAGCCATGCGTTAGTATGAGTGTGCTGAGCCTCCAGAC 120
DB 61 TCTTACGACAGAAAGCGCTTAGCCATGCGTTAGTATGAGTGTGCTGAGCCTCCAGAC 120

QY 121 CCCCCCTCCCGGAGAGCCATAGTGTGCTGCGGAACCGGTGAGTACACCCGAATTCGAG 180
DB 121 CCCCCCTCCCGGAGAGCCATAGTGTGCTGCGGAACCGGTGAGTACACCCGAATTCGAG 180

QY 181 GACGACCGGGTCCCTTCTTGATCAACCCGCTCAATGCTCGAGATTGGCGGTCGCCCC 240
DB 181 GACGACCGGGTCCCTTCTTGATCAACCCGCTCAATGCTCGAGATTGGCGGTCGCCCC 240

QY 241 GCGAGACTGTAGCCGAGTATGTGTTGGTTCGGAAGAGCCCTTGTGTAATCTGCTGATAG 300
DB 241 GCGAGACTGTAGCCGAGTATGTGTTGGTTCGGAAGAGCCCTTGTGTAATCTGCTGATAG 300

QY 301 GTGCTTGGAGTGCCTCCGAGAGTCTGCTAGACCGTGCACCACTGAGCAGCAATCCTAAAC 360
DB 301 GTGCTTGGAGTGCCTCCGAGAGTCTGCTAGACCGTGCACCACTGAGCAGCAATCCTAAAC 360

QY 361 CTCAAGAAAAACCAAGCTAACCAACCCGCGCCACAGAGAGCTCAAGTTCCCGGGCG 420
DB 361 CTCAAGAAAAACCAAGCTAACCAACCCGCGCCACAGAGAGCTCAAGTTCCCGGGCG 420

QY 421 GTGCTCAGATCGTGTGAGGTTTACCTGTGTCGCGGAGGGGCGCCAGGTTGGGTGTC 480
DB 421 GCGGTTCAGATCGTGTGAGGTTTACCTGTGTCGCGGAGGGGCGCCAGGTTGGGTGTC 480

QY 481 GCGGACTAGGAAGGCTTCCGAGCGGTGCGAACCTCGTGGAGGCGCAACCTATCCCAA 540
DB 481 GCGGAGCAGAGAGACTTCCGAGCGGTGCGAACCTCGGAGGTAGAGTCAAGCTATCCCAA 540

QY 541 AGGCTCGCCGACCCGAGGCGAGGCGCTGGGCTCAGCCCGGGTACCCCTTGGCCCTCTATG 600
DB 541 AGGCACGTCGCGCCGAGGCGAGGCGCTGGGCTCAGCCCGGGTACCCCTTGGCCCTCTATG 600

QY 601 GCAATGAGGCGCTGGGTTGGGAGAGATGGCTCTCTGTAACCGCGGCTCCGCGCTAGTT 660
DB 601 GCAATGAGGCGCTGGGTTGGGAGAGATGGCTCTCTGTAACCGCGGCTCCGCGCTAGTT 660

QY 661 GGGGCCCCACAGACCCCGGGGTAGGTGCGTAACTTGGGTAAAGTCAATCGATACCCCTTA 720
DB 661 GGGGCCCCACAGACCCCGGGGTAGGTGCGTAACTTGGGTAAAGTCAATCGATACCCCTTA 720

QY 721 CATGCGGCTTCCCGATCTCATGGGATACATTCGCTCGTCCGCGCCCCCTTGGGGGCG 780
DB 721 CATGCGGCTTCCCGATCTCATGGGATACATTCGCTCGTCCGCGCCCCCTTGGGGGCG 780

QY 781 CTGCGAGGCGCTTGGGAGAGAGTGTGCGGAGTGTGAGGAGAGAGTGTGAGGAGAGAG 840
DB 781 CTGCGAGGCGCTTGGGAGAGAGTGTGCGGAGTGTGAGGAGAGAGTGTGAGGAGAGAG 840

QY 841 GGAACTTGGCGGCTTGTCTTCTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 841 GGAACTTGGCGGCTTGTCTTCTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900

```

```

841 GGAACTTCTGCTGCTGCTTCTTCTATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
901 TCCAGCTTCCGCTTATAGAGTCCGCAACGCTGCCGGATATACCAATGTACGACAACTG 960
901 TGCCCGCTTCCGCTTACCAAGTCCGCAATCTCTGCGGCTTACCAATGTACCAATGAT 960

961 GCTCCAACTCAAGCATTTGTATGAGGCGAGGAGCTGATCATGATATCCCGGGTGG 1020
961 GCTCCAACTCAAGCATTTGTATGAGGCGAGGAGCTGATCATGATATCCCGGGTGG 1020

1021 TGCCCTGCTTCCGCTTACCAAGTCCGCAATCTCTGCGGCTTACCAATGTACCAATGAT 1080
1021 TGCCCTGCTTCCGCTTACCAAGTCCGCAATCTCTGCGGCTTACCAATGTACCAATGAT 1080

1081 CGCCAGGAAATCCAGCTTCCGCTTACCAATCTCTGCGGCTTACCAATGTACCAATGAT 1140
1081 CGCCAGGAAATCCAGCTTCCGCTTACCAATCTCTGCGGCTTACCAATGTACCAATGAT 1140

1141 GAGCGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
1141 GAGCGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200

1201 TCTCCAGCTGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
1201 TCTCCAGCTGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260

1261 CAATCTATCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
1261 CAATCTATCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320

1321 CACTCAACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
1321 CACTCAACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380

1381 TGCTGCGGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
1381 TGCTGCGGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440

1441 ACTGGGCTAAGGTTCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
1441 ACTGGGCTAAGGTTCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500

1501 CGAGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1501 CGAGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560

1561 TCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
1561 TCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620

1621 TGAATTCGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
1621 TGAATTCGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680

1681 TCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
1681 TCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740

1741 AGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
1741 AGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800

1801 GCGATTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
1801 GCGATTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860

1861 ATTGCTTCACTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
1861 ATTGCTTCACTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920

1921 ATAGCTGGGCGGAGATGAGACAGCTGATGCTCTCAACAAACAGCGCTGCTGCTGCTGCTGCT 1980
1921 ATAGCTGGGCGGAGATGAGACAGCTGATGCTCTCAACAAACAGCGCTGCTGCTGCTGCTGCT 1980

```


1981 GCACTGGTTCGGCTGTACATGGATAGTACTGGCTTCACTAAGACGTGGGAGTTC 2040
1981 GCAATTTGGTTTGGTTTACCTGGATGAATCAATGGATTCACCAAGTGTGGGAGGCG 2040
2041 CCCCGTGAACATCGGGGGGTGGTAAACCGACCTTGATCTGCCCAACGAGTCTCTCC 2100
2041 CCCCTTGTCTCATCGGAGGGTGGGCAACACACTTGTCTCTGCCCACTGATGTTC 2100
2101 GGAAGACCCCGAGGCTACTTACACAAATGTGGCTCGGGGCCCTGTGTGACACTAGGT 2160
2101 GCAAGCATTCGGAAGCCACATCTCTGGTGGGTCCGGTCCCTGGATTAACCCAGGT 2160
2161 GCTTAGTAGATACCCATACAGGCTTTGGCACTACCCCTGACCTCTCAATTTTTCCTCT 2220
2161 GCATGCTGACATACCCGCTATAGGCTTTGGCACTATCTTGTACCAATTAACCAAT 2220
2221 TTAAGTTAGGATATGTGGGGGCTGGAGACAGGCTCAATGCCCATGCAATTGGA 2280
2221 TCAAGTCAGGATGTACGTGGAGGGTGGAGCAGAGCTGGAAGCGGCTTCAACTGGA 2280
2281 CTCGAGGAGAGGCTGTAACTTTGGAGACAGGATAGTCAAGTCAAGCTCAGCTGCTGC 2340
2281 CGCGGGCGAAGCGTGTGATCTGGAGACAGGAGCTCGAGCTCAGCCCATTTGCTGC 2340
2341 TGTCTAACACAGAGTGGCAGATACGCTGTGCTTTTCAACACCTACCGGCTTTATCCA 2400
2341 TGTCCACACACAGTGGCAGGCTCTTCCGTGTTCTTTTCAACACCTTGCAGCTTTGCCA 2400
2401 CTGGTTTGCATCCATCCATCAGAACATCGTGGACGTGCAATPACCTGTACGGTGTAGGT 2460
2401 CGGGCTCATCCACCTCCACAGAACATTTGTGACGTGCAGTACCTTGTACGGGTAGGT 2460
2461 CAGCGTTTGTCTCTTTTGAATCAATGGAGTACATCTGTGCTTTTCTTCTCTCTGG 2520
2461 CAAGCATCGCTTCTGGGCCATTAAGTGGAGTACGTGTTCTCTGTTTCTCTCTGTTG 2520
2521 CAGACGCGGCTGTGCTGCTTGTGATGATCTGCTGATAGCTCAGGCTGAGGCGG 2580
2521 CAGACGCGGCTGTGCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 2580
2581 CTTTAGAACCTTGTGTTTCAATGGCGGCTCGTGGCGGAGCGATGATTTCTCT 2640
2581 CTTTGGAGAACCTGTAATCAATGAGATGCTGCTGGCGGAGCGACGAGCTTGTGT 2640
2641 CTTTCTGTTGTTCTGCTTGTGCTGATGATGATGATGATGATGATGATGATGATGATG 2700
2641 CTTTCTGTTGTTCTGCTTGTGCTGATGATGATGATGATGATGATGATGATGATGATG 2700
2701 GGTATGCTTTTATGCGGCTATGCGGCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 2760
2701 TCTAGGCTTCTAGGGATGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
2761 CTTAGGCTTGGACGGGAGATGGTGTGATGCTGGGGGTGGGTTCTTGTAGTCTGG 2820
2761 CATAGGCACTGGACACGGAGTGGCGGCTGCTGGGGGGGCTTGTCTTGTGGGTTAA 2820
2821 TATTTCTGACCTTGTACACATACAAAGTGTCTCTAGGCTCATATGGTGTAT 2880
2821 TGGGCTGACTCTGTGGCCATTTACAGCGCTACATCAGTGTGATGTTGGTGGCTTC 2880
2881 AATATCTTATCACCAGACGGAGGCGCATGCAAGTGTGGTCCCGGCTTCAACGTTTC 2940
2881 AGTATTTTCTGACCAAGTGTAGAGCGCACTCAGTGTGGTTCCTCCCTTCAACGTTTC 2940
2941 GGGAGGCGGCGATGCGATTCCTCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
2941 GGGGGGGGCGGATGCGCTTCTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 3000
3001 ACATCACCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060
3001 ACATCACCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060

3061 CGAGAGTGGCTACTTCTGTCGGCTCAAGGGCTCATTCGTGATGCAATGTAGTGGAA 3120
3061 TTAAGTCCCTTACTTGTGCGGTCAAGGCTTCTCCGATCTGCGCTAGCGGGA 3120
3121 AAGTCGCGGGGTCTATTATGTCCAAATGTCTTCAAGTGTGGGCGCTGACAGTA 3180
3121 AGATAGCGGAGGTCAATAGGTGCAATGGCCATCATCAAGTGTAGGCGCTTACTGGCA 3180
3181 GGTACGTTTATACCATCTTACCCCACTGCGGACCTGGGCGCCAGCGGCTTAGGAGCC 3240
3181 CCTATGTATACCACTCTCACCCTCTTCGAGACTGGGCGCAACGCGCTTGGAGATC 3240
3241 TTGCGGTGGGCTAGAGCCCGTCTCTCTCCGCAATGAGAGCAAGGTCTATCACCCTGG 3300
3241 TGGCGGTGCTGTGGAACCACTCTCTCTCCGCAATGAGAGCAAGGTCTATCACCCTGG 3300
3301 GAGCAGACACGCTGCTGTGGGACATCATCTTGGGTCTACCGTCTCCGCGGAGGG 3360
3301 GGGCAGATACCGCGCTGGGTGACATCATCAACGCTTGGCGCTCTGCGCGTAGGG 3360
3361 GGAAGGAGATATTTTGGGACCGGCTGATGTCTCGAAGGCGAAGGTGCGACTCCTTG 3420
3361 GCGGAGATATCTCTTGGGCGAGCGCGAATGGTCTCGAAGGTGGAGGTGCTTG 3420
3421 CGCCCATCAGGCTACTCTCCAAACGCGGGCGTACTTGGTTGATCATCATCAGTACC 3480
3421 CGCCCATCAGGCTACTCGCCAGCAGAGAGGCGCTCTAGGCTGTATATCACCAGCC 3480
3481 TCACAGCGCGGACAGAACACAGTGAAGGGGAGGTTCAGTGGTTCACCGCAACAC 3540
3481 TGACTGGCGGGAACAAACCAAGTGGAGGTGAGGTCCAGATGCTGTCACTGCTACCC 3540
3541 AATCTTCTTGGGCGCTGATCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3600
3541 AATCTTCTTGGGCGCTGATCAATGGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 3600
3601 CGAAGACCTTAGCGGCTCCAAAGTCCAAATCACCCAAATGTACCAATGTAGACCTGG 3660
3601 CGAGGACCATCGCATCACCCAGGCTCTGTCTATCCAGATGTATACCAATGTGACCAAG 3660
3661 ACCTGCTGGTGGCAGCGGCGCCCGGGGGCGCTCCATGACACCATGAGCTGTGGCA 3720
3661 ACCTTGTGGGTGGCGCGCTCTCAAGTGTCCCGCTCAATGACACCTGACCTGCGGCT 3720
3721 GCTCGGACCTTACTTGTGTCAGGACATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 3780
3721 CTTGCGACCTTACTGCTGTCAGGAGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3780
3781 ACAGCAGGGAGTCTACTCTCCCGCAGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCT 3840
3781 ATAGCAGGGTAGCTGCTTTCGCGCGGCGCAATTCCTCTCTCTCTCTCTCTCTCTCT 3840
3841 GTCCATGCTTGTGCTTTCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3900
3841 GTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3900
3901 GGGGGTTCGCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3960
3901 GTGAGTGGCTAAGCGGTGACTTTTATCTCTGTGGAGACCTTAGAGCAACCATAGAT 3960
3961 CTCGCTCTTTCACAGAACCTCAACCCCGGCTGTACCGGACACATTTCCAAAGTGGAC 4020
3961 CCGCGGTGTTTCACGCAAACTCTCTCTCAACAGCAGTGGCGGAGCTTCCAGTGGCG 4020
4021 ATCTGCGCTCTTACTGCGCAGCGGACAGCAGCAGGCTGCGGCTGCTGCTGCTGCTG 4080
4021 ACCTGCGCTCTTCTCCACCGGAGCGGTAGAGCAACAGGCTCCCGGCTGCTGCTGCTG 4080
4081 AAGGCTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4140
4081 AGGCTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4140
4141 ATATGTCAGGCGACAGGCTATACCGCTTACATCAGAACCTGGGGGTAGGACCATTACCA 4200

Db 4141 ACATGTCGAAGGCCATGGGGTGTGATCTTAATACAGGACCGGGGTGAGAACTATACCA 4200
QY 4201 CGGGCGGCTCCATTAGCTACTCACTCACTATGGCAAGTTCCTTGCAGCGGTGGGTGCTG 4260
Db 4201 CTGGCAGCCCACTCAGGTACTCCACTACGGCAAGTTCCTTGCAGCGGGGTGCTCAG 4260
QY 4261 GGGGCGCTATGACATCAATAATGTGATGAGTGCACACTCAATGTACTGACTPACCATCT 4320
Db 4261 GAGGTGCTATGACATAATATTTGTGACAGTGCACCTCCACGGATGCCACATCCATCT 4320
QY 4321 TGGGCTATGGCAAGTCTCTGGACCAAGCGAGAGAGCGGTGAGCGCGCTGCTGCTGCTG 4380
Db 4321 TGGGCTATGGCAAGTCTCTGGACCAAGAGAGAGCGGTGAGCGCGCTGCTGCTGCTG 4380
QY 4381 CCACCGCTACACCTCCGGGATCGGTACCGTGCACACCCCAATATCGAGGAATAGGCC 4440
Db 4381 CCACTGCTACCTCCGGGCTCGTCACTGTGTCCATCTCAATCGAGGAGGTGCTC 4440
QY 4441 TGTCACCAATGAGAGATCCCTTCTATGSCAAAGCCATCCCATGTAGAGCCATCAAGG 4500
Db 4441 TGTCACCAAGAGATCCCTTCTATGSCAAAGCCATCCCATGTAGAGGTGATCAAGG 4500
QY 4501 GGGGAGGATCTCATTTCTGCGATTCGCAAGTGAAGAAATGTGAGAGCTCGCGCAAGC 4560
Db 4501 GGGGAGAGATCTCATTTCTGCGACTCAAGAGAGAGTGCAGAGCTCGCGCGAAGC 4560
QY 4561 TGACAGGCTCGGACTGAAGCGTGTAGCATATPACCGGGCCCTGTATGTCTCGTCAATC 4620
Db 4561 TGTCGCAATGGGATCATGATCGGTGCGCTACTACCGCGTCTTGACGTCTGTGATCC 4620
QY 4621 CGCTATCGGAGAGTGTGTCTGTGCAACAGACGCTCTAATGACGGGTTCAACGGGG 4680
Db 4621 CCAACAGCGCGATGTGTCTGTGCGACCGATGCTCTCATGACTGTGCTTACCGGG 4680
QY 4681 ATTTTGACTCAGTATGCACTGCAATACATGTGTCAACAGAGAGTGTGAGCTTGTG 4740
Db 4681 ACTTCGACTCTGTGATAGACTGCAACAGTGTGTCTCACTCAGACAGTGTGATTT 4740
QY 4741 ATCCACCTTACCATTTGAGACGACGACCGTGCACCAAGCGGGTGTGCGCTCGCAAC 4800
Db 4741 ACCCTACCTTTACATTTGAGACACACGCTGCCCGTCCCGAGATGTGTCTCAGAGCTCAAC 4800
QY 4801 GGGAGGTAGACTGGCAGGGGTAGGAGTGGCACTACAGTTTGTGACTTCCAGGAGAAC 4860
Db 4801 GCGGGGACGAGCTGGCAGGGGAGGACGAGCATCTACAGATTTGTGCAACCGGGGAGC 4860
QY 4861 GGCCTCGGCACTGTGATTTCTCGGTCTGTGTGAGTGTATGACGCGGGCTGTGCTT 4920
Db 4861 GCGGCTCGGCACTGTGATTTCTCGGTCTGTGTGAGTGTATGACGCGGGCTGTGCTT 4920
QY 4921 GGTATGAGCTCAGCGCGCTGAGACCTCGGTGTAGTTGCGGGCTTACTTAATACACAG 4980
Db 4921 GGTATGAGCTCAGCGCGCTGAGACCTAGTTAGTTGCGGGCTTACTTAATACACAG 4980
QY 4981 GGTTCGCGCTGCGCAGCACTCTGAGTGTGGAGAGGCTTTCAAGGCTTACCC 5040
Db 4981 GGTTCGCGCTGCGCAGCACTCTGAGTGTGGAGAGGCTTTCAAGGCTTACCC 5040
QY 5041 ACATAGATGCCACTTCTGTGCCAGACTAAACAGCGAGGACAACTTTTCTTACCTGG 5100
Db 5041 ATATAGATGCCACTTCTGTATCCAGACAAAGAGAGTGGGAGAACTTTCTTACCTGG 5100
QY 5101 TGGCATATCAAGCTACAGTGTGCGCCAGGGCTCAAGCTCCACTTCCATCTGTGGACCAAA 5160
Db 5101 TAGCGTACCAAGCAACGCTGTGCGCTAGGGCTCAGCGCTCCCGCTTCTTACCTGG 5160
QY 5161 TGTGGAAGTGTCTCATAGGCTGAAACCTACACTGCAAGGCGCAACACCGCTGTGTATA 5220
Db 5161 TGTGGAAGTGTGTGATCGCGCTTAAACCCAGCTTCCATGGGCGCAACACCGCTGTATA 5220
QY 5221 GGCTAGGAGCGCTCCAAATGAGGTCTATCTTCAACACCGCTTAAATATGATATGG 5280

Db 5221 GACTGGCGCTGTTTCAGAATGAGTCAACCTGACGCAACCAATACCAAAATACATCA 5280
QY 5281 CATGATGTGGCTGACCTGAGGTGCTCACTAGCACTGGGTGCTGGTAGCGAGTCC 5340
Db 5281 CATGATGTGGCTGACCTGAGGTGCTCACTAGCACTGGGTGCTGGTAGCGAGTCC 5340
QY 5341 TTGCAAGTTTGGCCCGATCTGCTCTGACGACAGGAGTGTGCTCATTTGTGGCAGATCA 5400
Db 5341 TGGCTGCTTGTGGCCCGATTTGCTCTCAACAGGCTGCTGCTGCTATGTTGGGCAAGATG 5400
QY 5401 TCTTCTCGGGGAGCGAGTGTGCTTCCGACAGGGAAGTCTCTTACAGAGTTCGATG 5460
Db 5401 TCTTCTCGGGGAGCGAGTGTGCTTACCTGACAGGAGTTCCTTACAGAGTTCGATG 5460
QY 5461 AGATGGAAGAGTGTGCTCTCAAACTTCTTATCATGAGAGGGAATGAGCTTCGCGAGC 5520
Db 5461 AGATGGAAGAGTGTGCTCTCAAACTTCTTATCATGAGAGGGAATGAGCTTCGCGAGC 5520
QY 5521 AATTTAAGCAAAAGGGCTCGGGTTGTTGCAACCGGCCACCAAGCAAGGGAGGCTGCTG 5580
Db 5521 AGTTCAAGCAGAGAGGCTTCGGCTCTCTGACAGCGCGTCCCGCAAGCAGAGTTATCA 5580
QY 5581 CTCCGCTGTGAGTCCAAAGTGGCGAGCCCTTGAGACCTTCTGGCGGAAGACATGTGA 5640
Db 5581 CCCTGCTGTCCAGACCAACTGCGAGAACTCGAGGTCTTCTGGCGGAGACATGTGA 5640
QY 5641 ATTTTCATCAGCGGAATPACAGTACCTAGCAGGCTTATCCACTCTGCTGGAACCCCGCA 5700
Db 5641 ATTTTCATCAGTGGATACAACTTTGGGGGCTCTCAACGCTGCTGTTAAACCCCGCA 5700
QY 5701 TAGCATCATGTGCACTTACAGCTTCTATCACTAGCCGCTCACCACCAACCAACCC 5760
Db 5701 TTGCTTCACTGTGCTTTTACAGTGTGCTTACAGTGTGCTTACAGTGTGCTTAAACCC 5760
QY 5761 TCTGCTTTAATCTTGGGGGATGGGTGGCTGCGCAACTGCTCTCTCCAGCGTGTGCT 5820
Db 5761 TCTCTTTCAATTTGGGGGATGGGTGGCTGCGCAACTGCTCTCTCCAGCGTGTGCT 5820
QY 5821 CAGCTTTCTGCGCGCGCTGCTGCGCGAGCGGCTTGGCAGCATAGGCTTTGGGAAG 5880
Db 5821 CGGCTTTCTGCGCGCGCTGCTGCGCGAGCGGCTTGGCAGCATAGGCTTTGGGAAG 5880
QY 5881 TGTCTGTGACATCTTGGCGGCTATGGGCGAGGGGTAGCGCGGCACTGCTGGCTTTTA 5940
Db 5881 TGTCTGTGACATCTTGGCGGCTATGGGCGAGGGGTAGCGCGGCACTGCTGGCTTTTA 5940
QY 5941 AGTCTATGAGCGCGAGGTGCTTCCACGAGACCTGCTGCACTTACTTCTTCCCTG 6000
Db 5941 AGATCATGAGCGGTGAGGTCTTCCACGAGACCTGCTGCACTGCTGCTGCGCATCC 6000
QY 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6060
Db 6001 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6060
QY 6061 GCGCGGAGAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6120
Db 6061 GCGCGGAGAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6120
QY 6121 ACCACGCTCTCCCTTACGCACTATGTGCTTGGAGGCGACGCTGCGACGATGCTGCT 6180
Db 6121 ACCATGTTTCCCGCAGCACTAGCTGCTGCGAGAGCGATGCAAGCGCGCGCTGCTG 6180
QY 6181 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6240
Db 6181 TACTCAGAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6240
QY 6241 ACTGCTCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6300
Db 6241 AGTGTACCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6300
QY 6301 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6360
Db 6301 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6360

QY 6361 CTTTCTCTGTCATGCAACGCGGGTACAAAGGAGTCTGCGGGGGACGGCATCATGCAAA 6420
Db |||||
QY 6361 CTTTGTGTCTCTGCGACGCGGGTATAGGGGGTCTGCGGAGGAGACGGCAATTATGCACA 6420
Db |||||
QY 6421 CCACTGCGCCATGCGGAGCAGATGCGCGGACATGTCAAAAACGGTTCCATGAGGATCG 6480
Db |||||
QY 6421 CTCGTGCACTGTGGAGTCTGAGATCACTGGACATGTCAAAAACGGGACGATGAGGATCG 6480
Db |||||
QY 6481 TAGGSCCTPAGAACCTGCGAGCAACACGTGGCGCGGAACGTTCCCATCAACGACATACACA 6540
Db |||||
QY 6481 TCGTCTTAGACCTGCGAGAACATGTGAGTGGAGCGTTCCCATTAACGCTTACACA 6540
QY 6541 CCGGACCTTGCACACCTTCCCGCGGCCCAACTAFTCCAGGGCGCTATGCGGGTGGCTG 6600
Db |||||
QY 6541 CCGGCCCTGTACTCCCTCTCTGCGCGCACTATAAGTTCGCGCTGTGGAGGGTGTCTG 6600
QY 6601 CTGAGGATAGCTGAGAGTTAGCGGTGCGGGATTTCCATPCTGAGCGGCATGACCA 6660
Db |||||
QY 6601 CAGAGGAATACGTGGAGATAAGCGGGTGGGGACTTCCATACGTATCGGTATGACTA 6660
QY 6661 CTGACAAAGTAAAGTCCATGCGCAGGTTCCGGCCCCCGAATTTCTACGGAGGTGGATG 6720
Db |||||
QY 6661 CTGACAACTTAATGCGCGTGCAGATCCCATCGCCCCGAATTTTTCAGAAATTGGAGC 6720
QY 6721 GAGTGGGTGGACAGGTACGTCCGGCGTGCAGAACCTTCTTACGGGAGACGFTACGT 6780
Db |||||
QY 6721 GGGTGGCGCTACATAGGTTTGGCGCCCTTGCAGGCCCTTGTGCGGGAGGAGGTATCAT 6780
QY 6791 TCCAGTGGCGCTCAACAAATCTGCTGGGTGCGAGTCCCATCGAGCGCGACCGG 6840
Db |||||
QY 6791 TCAGAGTAGGACTCCACAGATACCGGGTGGGTGCGCAATTAACCTTGGAGCGCGAACCGG 6840
QY 6841 ACCTAACAGTCTTACTTCCATGCTCAGCATCCCTCCCAANTACAGACAGACGCGCTA 6900
Db |||||
QY 6841 ACCTAGCGGTGTGAGTCTCATGCTCATGATCCCTCCCATATAACAGAGGCGCGCG 6900
QY 6901 AGCTPAGCTGGCTAGAGGTTCCCGCTCTTTAGCCAGGTCACTAGCTAGCCAGTTGT 6960
Db |||||
QY 6901 GGAGAAGTTGGCGAGAGGTCACCCCTTCTATGGCCAGCTCTCGGCCAGCCAGCTGT 6960
QY 6961 CTGCGCTCTTTGAGGGGACATGCACTACCCACATGACTCCCGGACGCTGACTCA 7020
Db |||||
QY 6961 CCGCTCCATCTCTCAAGGAACTTGCACGCGCAACCATGACTCCCTCAGCGCGAGCTCA 7020
QY 7021 TCGAGGCCAACTCTGTGGCGGCGAGAGATGGCGGAAACATCACTCGCTGGAGTCAG 7080
Db |||||
QY 7021 TAGAGGCTAACTCTGTGGAGGAGAGATGGCGGCAACATCACCGGTTGAGTCAG 7080
QY 7081 AGAATAAGGTAGTAATCTTGGACTCTTTGAAACCGTTTCAACCGGTTCAACGAGGGGATGAGGG 7140
Db |||||
QY 7081 AGAACAAAGTGGTGATTTCTGGACTCTTTCGATCCCGCTTGTGGCAGAGGAGATGAGCGG 7140
QY 7141 AGATATCCGTGCGCGGAGATCTTGGGAAATCAGGAAGTTCCCTCAGCGTTGCCCA 7200
Db |||||
QY 7141 AGTCTCTGATCCCGACGAAATCTTGGGAACTCTCGGAGATTCGCGGGCCCTGCCCG 7200
QY 7201 TATGGGCGCGCGGAGTACATCTCCACTGCTAGAGTCTTGGAGGACCGCGGACTAG 7260
Db |||||
QY 7201 TTTGGGCGCGCGGACTACAAACCCCGCTAGTAGAGACGTGGAAAGCCTGACTAG 7260
QY 7261 TCCCTCCGTGTACAGGATGCCATTCGCACTTACCAAGGCTCTCAATACCACTC 7320
Db |||||
QY 7261 AACCACTGTGTGTCATGGCTCCCGCTACCACTTCCAGCGTCCCTCTGTGCTCCG 7320
QY 7321 CACGAGAAAGAGGAGGTGTCTTGACAGAAATCCAAATGTGTCTTGTCTGGCGGAGC 7380
Db |||||
QY 7321 CTCGGAAMAGGCTAGGTGTCTTCAACCGAATCAACCTTATCTACTGCTTGGCGGAGC 7380
QY 7381 TCGCCACTAAGACCTTCGGTAGCTCCGGATCGTCGCGCTTGAATAGCGGACGCGGACCG 7440
Db |||||
QY 7381 TTGCCACCAAAAGTTTGGACGCTCTCTCACTTCGGGCAATTACGGCGGACATACGACA 7440

QY 7441 CCCTTCTGACCTGGCTCCGACGACGGTGACAAAGGATCCGACGTTGAGTCGTACTCCT 7500
Db |||||
QY 7441 CATCTCTGAGCGCGCTTCTGGCTGCCCCCGACCTCGACGTTGAGTCCTATTCTT 7500
QY 7501 CCATGCCCCCTTGAAGGGAGCGCGGGGACCCCGATCTCAGGACGGGTCCTGGCTA 7560
Db |||||
QY 7501 CCATGCCCCCTTGAAGGGGAGCCTGGGATCCGGATCTCAGGACGGGTCATGGTCGA 7560
QY 7561 CCGT---GAGTAGGAGGCTAGTGAGGATGTCTGCTCTGCTCAATCTCCTATACGTGGA 7617
Db |||||
QY 7561 CGGTCACTAGTGGGCGGACACGAGAGATGTCTGTGCTGCTCAATCTCTTATCTCTGGA 7620
QY 7618 CAGCGCCCTGATCACGCTATGCGCTCGGAGGAAATAGCTGCCCATCAACCCGTGA 7677
Db |||||
QY 7621 CAGCGCACTCTCACCCCTGCGCTCGGAAAGCAAAACCTGCCATCAACGCACTGA 7680
QY 7678 GCAACTCTTGTGCTGCTCACCAACATGCTACGCCCAACAATCCCGGAGCGCAAGCC 7737
Db |||||
QY 7681 GCAACTCTTGTGCTAGCGCATCAATCTGTGTATTCACCATCTTCAGGAGTCTTGCC 7740
QY 7738 TCCGCGAAGAAAGGTCACTTTGACAGATTGCAAGTCTCTGGAATCATTAACGGGAGC 7797
Db |||||
QY 7741 AAGGCGAGAAAGTCACTTTGACAGACTGCAAGTCTTGGACAGCCATTAACGAGGAGC 7800
QY 7798 TACTCAAGGAGTGAAGCGGAGCGCTCCACAGTTAAGGCTPAGCTTCTATCTATAGAG 7857
Db |||||
QY 7801 TGCTCAAGGAGTCAAAAGCAGCGCTCAAAAGTGAAGGCTAACTTCTATCCGTAGAGG 7860
QY 7858 AGGCTCAGAGTCAAGCGCCCCCATTGCGGCAATCCAAATTTGGCTATGCGGCAAGG 7917
Db |||||
QY 7861 AAGTTGAGCTGAGCGCCCCCATTTACGCCAATTCRAAGTTTGGCTATGCGGCAAG 7920
QY 7918 ACCTCGGAACTTATCCAGCAGGCGCCCTTAACCACTCCGCTCCGTGTGGAGGACTTGC 7977
Db |||||
QY 7921 ACCTCGGTGCTCATCCAGAAAGCGCTAGCCACATCACTCGTGTGGAAGACCTTC 7980
QY 7978 TGGAGACACTGAACCAATTGACACACCATCATGGCAAAAGTGAGGTTTTCTGCG 8037
Db |||||
QY 7981 TGGAGACAGTGTAAACCAATAGACACTACCATATGGCCAAAGACGAGTTTTCTGCG 8040
QY 8038 TCCAAACAGAGAGGAGCGCGCAAGCCAGCTCCCTTATCGTATCCAGACCTGGAG 8097
Db |||||
QY 8041 TTCAGCTTGAAGCGGGTCTGAAGCAGCTCTCTCATCTGTCTCCCGACTGGCG 8100
QY 8098 TTGCTGTATGCGAGAGTGGCCCTTTAAGAGTGGTCTCCACCTTCTCTCAGCGCTGA 8157
Db |||||
QY 8101 TGCGCTGTGCGAGAGATGCCCTGTAGGACGTTGTAGCAAGCTCCCTCGCCGCTGA 8160
QY 8158 TGGCTCTCATACGATTTCAATTAATCTCCCAAGCAGCGGTCGAGTCTCTGTGATA 8217
Db |||||
QY 8161 TGGAAAGCTCTTACGGATTCATTAACCAAGCAGCGGTTGAATTCCTCTGCAAG 8220
QY 8218 CTTGAAATCAAGAAATGCTTGGCTTCTCATATGACACCCGCTGTTTGACTCAA 8277
Db |||||
QY 8221 CTTGAGTTCAGAGACCCCGATGGGTTCTGTATGATACCCGCTGTTTGACTCCA 8280
QY 8278 CCGTCACTGAGAGTCAATCTGTGTGAGGAGTCAATTAACCAATGTGTGACTTGGCC 8337
Db |||||
QY 8281 CAGTCACTGAGAGCAATCTGTACGAGGAGGAGCAATTAACCAATGTGTGACTGGACC 8340
QY 8338 CCGAGGCGCAGACGACATAGGTGCTCAGAGCGGCTTATACATGCGGGTCCCTGA 8397
Db |||||
QY 8341 CCCAGCGCGCTGGCCATCAAGTCCCTCACTGAGAGGCTTATGTTGGGGGCCCTCTTA 8400
QY 8398 CTACTCAAAAGGCGAGAACTGCGGTTATCCCGCTGCGCGCAAGTGGCGCTGTGACGA 8457
Db |||||
QY 8401 CCAATTAAGGGGGGAAACTGCGGCTACCGAGTCCCGCGCGAGCGGCTGTGACAA 8460
QY 8458 CTAGTCCGCTAATACCTTCACTTGTGAGAGGCACTTCAGAGCTGTGAGCTGCA 8517
Db |||||
QY 8461 CTAGTGTGGTAAACACCTCTACTTGTACATCAAGGCGCGGCGAGCTGTGAGCGGAG 8520
QY 8518 AGCTCCAGGACTGCAAGTGTCTGTGAACGAGAGACCTTGTCTGTATCTGTGAAGCG 8577

8521	Db	GGCTTCAGGACTGCACCATGCTCGTGTGTGCGGAGCACTTAGTCGTTATCTGTGAAAGTG	8580
8578	Qy	CGGAAOCCAGGAGGATGCGGCGGCCCTTAAGAGCTTCAOAGGCGTATGACTAGGTATT	8637
8581	Db	CGGGGTCCAGGAGGACGGGCGGAGCCTGAGAGCTTCAOAGGCGTATGACCAAGTACT	8640
8638	Qy	CCGCCCCCGGGGATCCGCCCAACAGAAATAGACCTTGAGCTGATAACATCATGTT	8697
8641	Db	CCGCCCCCGGGGACCCCCCAACAGAAATAGACTTGGAGCTTATACATCATGCT	8700
8698	Qy	CCTCAATGTGTCACTGCGGCAGATGCATCTGGCAAAAGGATATACCTCACCCGTG	8757
8701	Db	CCTCAACGTGTCACTGCCCCACAGCGCGCTGGAAAGAGGCTCTACTAOCCTAACCCGTG	8760
8758	Qy	ACCCACACCCCCCTTGCACGGGCTGGTGGGAGACAGCTAGACACACTCCAATCAACT	8817
8761	Db	ACCTTAACACCCCTCCGGAGACCGGTGGGAGACAGCAAGACACACTCCAGTCAATT	8820
8818	Qy	CTTGGCTAGGCAATATCATCATGATGCGGCCACCCCTATGGCAAGGATGATTCGTATGA	8877
8821	Db	CCTGGCTAGGCAACATTAATCATGTTTGCCCCCACTGTGGCGAGATGATCTGATGA	8880
8878	Qy	CTCACTTTTCTCCATCCTCTCTAGCTCAAGAGCAACTGTGAAAAAGCCCTGGATGTCAGA	8937
8881	Db	CCCATTTCTTAGCGTCTCATAGCCAGGATCAGCTTGAACAGCTCTTAACGTGAGA	8940
8938	Qy	TCTACGGGGCTGTACTCCATTGAGCACCTTGACCTACCTCAGATCATTTGAACGATCC	8997
8941	Db	TCTACGGAGCTGTACTCCATGAACACCTGGATCTACCTCCMATCATTTCAAGACTCC	9000
8998	Qy	ATGCTTTAGCGATTTACACTCCACAGTTACTCTCCAGGTGAGATCAATAGGGTCGCTT	9057
9001	Db	ATGCTCTCAGCGCATTTTCACTCCACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCG	9060
9058	Qy	CATGCTCAGGAACCTTGGGTACACCCCTGCGAACTCGAGACATCGGGCAGAGTG	9117
9061	Db	CATGCTCAGAAACTTGGGGTCCGGCTTGGCAGCTTGGAGACACCGGGCCGGAGCG	9120
9118	Qy	TCCGCGTAAGCTACTCTCCAGGGGGAGGGCGCCACTGTGGCAGATACCTCTTTA	9177
9121	Db	TCCGCGTAGGCTTCTCTCAGAGGAGCAGGGCTGCCATATGTGGCAAGTACCTCTTCA	9180
9178	Qy	ACTGGGAGTAAAGAACCAAGCTTAACTCACTCCATCCGGCGCGCGCTCCAGCTGAACT	9237
9181	Db	ACTGGGAGTAAGAACCAAGCTCAAACTCACTCCATAGGGCGCGCTGGCGGTGAACT	9240
9238	Qy	TGTCTGCTGTTCTCGTCTGCTTACAGCGGGGAGACATATACAGACTGTCTCGTG	9297
9241	Db	TGTCGGTGTGTTCAOGECTGGCTACAGCGGGGAGACATTTATCAGGCTGTCTCATG	9300
9298	Qy	CCGACCCCGCTGGTTTCCGTTGTGCTACTCTTACTCTCTGTAGGGGTAGGCATTTACC	9357
9301	Db	CCGGCCCCCGCTGGTTCTGTGTTTGTGCTACTCTCTGCTGCGCTGACAGGGGTAGGCATTTACC	9360
9358	Qy	TGCTCCCAACCGATGAACGGGAGCTAACACTCCAGGCCTTAAGCCATTTCTCTGTTT	9416
9361	Db	TCCTCCCAACCGAAGAAGTTGGGTAAACACTCCGGCTCTTTAGGCCATTTCTCTGTTT	9420
9417	Qy	TTTGTT	9476
9421	Db	TT	9480
9477	Qy	TT-----	9489
9481	Db	TTTTTTCTTT	9540
9490	Qy	TCCTTAATGTTGGTCCCATCTTAGCCCTTAGTCAOAGCTAGCTGTAAGGTCCGTGAGCC	9549
9541	Db	TCCTTAATGTTGGTCCCATCTTAGCCCTTAGTCAOAGCTAGCTGTAAGGTCCGTGAGCC	9600
9550	Qy	GCATGACTGCAGAGAGTGTGTGATCACTGGCCCTCTCTGCAATCATGT	9595

Db 2821 TGGCGTGAATCTCTGCGCAATTACAGCGCTATATCAGCTGTGATGTTGGGCTC 2880
Qy 2881 AATACCTTTATCAGAGCGGAGCGGACATGCAAGTGTGGTCCCGCCCTCAACCTTC 2940
Db 2881 AGTATTTTCAGCAGAGTAAAGCGCACTGCAAGTGTGGTTCGCCCTCAACCTTC 2940
Qy 2941 GGGAGGCGCGATGCCATATCTCTCTCAGCTGTGGGTTTCACTCAGAGTTAATTTTGG 3000
Db 2941 GGGGGGCGCGATGCGCTCATCTTACTCATGTGTAGTATACACCGACCTGTATTTG 3000
Qy 3001 ACATCAACAATCTCTGTGTGCTATCTGCGCCGCTCAAGTGTCTCAGCTGCGATAA 3060
Db 3001 ACATCAACAATCTCTGTGCTATCTCTGCGCCATCTTGGACCGCTTGGATTTCTCAAGCCAGTTTGG 3060
Qy 3061 CGAGAGTCCGCTACTCTGTGCGCTCAAGGCTCATCTGTCGATGATGTAGTGGAA 3120
Db 3061 TTAAGTCCCTACTCTGTGCGCTTCAAGGCTTCTCGGATCTGCGGCTAGCGGA 3120
Qy 3121 AAGTCCGCGGGGTCAATATGTCAAATGGTCTTCAAGCTGGGCGCTGACAGTA 3180
Db 3121 AGATAGCGGAGGTCAATTAAGTCAATGGCTATCAATAGTTAGGCGCTTACTGCA 3180
Qy 3181 CGTAGTTTAAACATCTTACCGACTGCGGACTGGGCCACCGCGGCTTACAGACC 3240
Db 3181 CCTATGTGTATAACCATCTCAACCTCTCTGAGACTGGCGCAACACGCGCTGGAGATC 3240
Qy 3241 TTGGGTGCGGTGAGCGCTGCTCTCTCGGCTATGAGACCAAGTCTATCACTCTGG 3300
Db 3241 TGGCGGTGCTGTGAACCATCTGCTCTCTCGGATGAGACCAAGTCTATCACTCTGG 3300
Qy 3301 GAGCAGACACCGCTGCTGCGGACATCATCTTGGGTCTACCGCTCTCCCGCCGAAAGG 3360
Db 3301 GGGCAGATACCGCGCTGCTGCTGACATCATCAACGGCTTCCCGCTCTCTCCCGCTAGG 3360
Qy 3361 GGAAGCAGATTTTGGGACCGCTGATCTCTGAGGCGAAGGTGGCGATCTCTTG 3420
Db 3361 GCCAGGAGATCTGCTTGGGCGAGCGGAGTCTCCAGGAGTGGGTTGCTGCG 3420
Qy 3421 CGCCCATCAGCGCTTCTCCCAAACAGCGGGGTACTTGTGCTATCATCTAGCTAGCC 3480
Db 3421 CGCCCATCAGCGGTATCGCCAGCAGACAGAGCTCTCTAGGTTGTATATCAACAGCC 3480
Qy 3481 TCACAGCGCGGACAGAAACAGGTCGAGGAGGTTCAAGTGTCTTCTACCGCAACAC 3540
Db 3481 TGAATGGCGCGGACAAACAGTGGAGGTGAGTCCAGATCTGTCAACTGCTACCC 3540
Qy 3541 AATCTTCTCTGCGACCTGCATCAACGGGTGTGCTGACTGTCTACCATGGCGTGGCT 3600
Db 3541 AAACCTTCTGCGAACGTGCAATGGGTTGCTGACTGTCTACCGGGGCGGAA 3600
Qy 3601 CGAGACCTTAGCGGTCCAAAGTCCAAATCACCCAAATGTACACCAATGTAGACCTGG 3660
Db 3601 CGAGACCATCGCATCACCAAGGTCTCTGTCATCCAGATGTATACCAATGTGACCAAG 3660
Qy 3661 ACTCTGTGCGTGGCAGCGCGCCCGGGCGGCTCCATGACACCAATGTGAGCTGTGCA 3720
Db 3661 ACTTGTGGGTGGCGCTCTCAAGGTTCGCGCTCATTCACACCTGTACCTTGGGCT 3720
Qy 3721 GCTCGGACTTTACTTGTGTGAGACATGTGATGTCTATCCGTTGCGCGGAGGCG 3780
Db 3721 CCTCGGACTTTACTTGTGTGAGACATGTGATGTCTATCCGTTGCGCGGAGGCTG 3780
Qy 3781 ACAGCAGGGAAGTCTACTTCCCGCCAGCGCGCTCTCTACTGAAAGGCTCTCTCGGGT 3840
Db 3781 ATAGCAGGGTAGCTGTCTTCCCGCCCGGCAATTTCTTCTTGAAGGCTCTCTCGGG 3840
Qy 3841 GTCCATTTGCTTTGCTTGGGCGACGTGTGTGGGCTCTCTCGGCTGCTGTGTGACCC 3900
Db 3841 GTCCGCTGTGTGCGCGCGGACACGCGCTGGGCTTATCAGGCGCGGCTGTGTGACCC 3900
Qy 3901 GGGGGGTGCGAGAGCGGTGACCTTCATACCGTTGAGTCTATGGAATCTACCTACCGGT 3960
Db 3901 GGCTTCCGTTGCGCAGGACCATCTTTGAAATTTTGGGAGGCGCTCTTTACGGGCTCACTC 3960

Db 3901 GTGAGTGGCTAAAGCGTGGACTTTTATCCCTGTGAGAACCTTAGGACAAACATGAGAT 3960
Qy 3961 CTGCGTCTTCAGAGCAACTCAACCCCGCTGTACCGAGACATTCACAGTGGCAC 4020
Db 3961 CCCGGTTCACGGAACACTCTCTCCACAGAGTGGCCCGAGAGCTTCAGGTGGCCC 4020
Qy 4021 ATCTCACGCTCTCTACTTGGCAGCGCAAGAGCACCAAAAGTGGCGGTGTGTATGACGCC 4080
Db 4021 ACCTGATGCTCCACCGGCGAGCGTAAAGAGCACCAAGTCCCGGTGCGTACGACGCC 4080
Qy 4081 AAGGTACAGGTGCTGCTCTGAAACCGTCTGTCGGCCACCTTAGGTTTGGGGCGT 4140
Db 4081 AAGGTACAGGTGCTGCTCTGAAACCGTCTGTCGGCCACCTTAGGTTTGGGGCGT 4140
Qy 4141 ATATGTCAAAGGCACAGCGTATCGACCTAAATCAGAACTGGGGTAAGACCAATACCA 4200
Db 4141 ACATGTCAAAGGCACAGCGTATCGACCTAAATCAGAACTGGGGTAAGACCAATACCA 4200
Qy 4201 CGGCGGCTCCATTAAGTACTCCACCTATGGCAAGTTCCTTGGCGAGTGGGTGTCTG 4260
Db 4201 CTGGCAGCCCATCACTACTCCACCTTAGCGCAAGTTCCTTGGCGAGTGGGTGTCTG 4260
Qy 4261 GGGCGGCTATGACATCAATATATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Db 4261 GAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Qy 4321 TGGGATCGGCACAGTCTCGACCAAGCGGAGAGCGGTGGAGCGCGCTCTGCTGCTG 4380
Db 4321 TGGGATCGGCACAGTCTCGACCAAGCGGAGAGCGGTGGAGCGCGCTCTGCTGCTG 4380
Qy 4381 CCAACCGCTACCTTCGGGATTCGCTGTCACACCCCAATATCGAGAAATAGGCC 4440
Db 4381 CCACTGATCCCTTCGGGTCGCTCACTGTGTCTCCATCTCAATCACTCGAGAGTGTCT 4440
Qy 4441 TGTCACCAATGAGAGATCCCTCTTATGGCAAGCCATCCCATTTAGGCCCATCAAGG 4500
Db 4441 TGTCACCAATGAGAGATCCCTCTTATGGCAAGCCATCCCATTTAGGCCCATCAAGG 4500
Qy 4501 GGGGAGGCTATCATTTTCTGCAATTCACAAAGAAATGTGACGAGTCCCGCCCAAGC 4560
Db 4501 GGGGAGGCTATCATTTTCTGCAATTCACAAAGAAATGTGACGAGTCCCGCCCAAGC 4560
Qy 4561 TGACAGGCTTCGAGTGAACGCTGTAGTATTAACCGGGCTTGTGTGTGTGTGTGTGTGT 4620
Db 4561 TGCTGT 4620
Qy 4621 CGCTATCGGAGAGT 4680
Db 4621 CGACAGCGGATGT 4680
Qy 4681 ATTTTGAATCAGTATCAGTGTCAATATCATGTGTCAACAGAGAGTGTGTGTGTGTGTGT 4740
Db 4681 ACTTGAATCAGTATCAGTGTCAATATCATGTGTCAACAGAGAGTGTGTGTGTGTGTGT 4740
Qy 4741 ATCCACCTTCAACATGAGACAGACCGTGTCCCAAGCGGTGTGTGTGTGTGTGTGTGTGT 4800
Db 4741 ACCCTACCTTCAACATGAGACAGACCGTGTCCCAAGCGGTGTGTGTGTGTGTGTGTGT 4800
Qy 4801 GGGGAGTGAAGCTTGGCAGGGGTAGAGTGGCATCTACAGTTTGTGTGTGTGTGTGTGTGTGT 4860
Db 4801 GCGCGGAGAGCTGGCAGGGGTAGAGTGGCATCTACAGTTTGTGTGTGTGTGTGTGTGTGT 4860
Qy 4861 GGGCTCGGCACTGT 4920
Db 4861 GCGCTCGGCACTGT 4920
Qy 4921 GGTATGAGTCAAGCGCTGAGACCTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980
Db 4921 GGTATGAGTCAAGCGCTGAGACCTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980
Qy 4981 GGTGTGCGCTGTGCGAGGACCTGTGAGTGTGGGAGAGGCTTTCACAGGCTCAACCC 5040
Db 4981 GGCTTCCGTTGCGCAGGACCATCTTTGAAATTTTGGGAGGCGCTCTTTACGGGCTCACTC 5040

QY 1681 TCACCTCGTCCGGTCCCGAGCCATGGCCAGCTGGCCGCCATTCGCTGGTTCGCC 1740
Db 1681 TCACCTCTCAGGCTGTCTGAGAGTTGGCCAGCTGGCCAGCTTACCGATTTGGCC 1740
QY 1741 AGGGTGGGGCCCATCACTATATAAGCTTAACAGCTCGGATCAGAGGCTTATGCT 1800
Db 1741 AGGGTGGGGTCTTATCAGTTATGCAACGGAAGCGGCTCGACGAAAGCCCTACTGCT 1800
QY 1801 GCAATACGGCTCGACGCTGGTGGTGGTACCGCGCTCGAGGTGGTGGTTCAGTGT 1860
Db 1801 GGCATACGCTCGAAGCTTGGCAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1860
QY 1861 ATTGTTTCAACCCAGACCTGGTGGTGGGACACACGATCGTTCCGCTGGTTCCTAGT 1920
Db 1861 ATTGCTTCACTCCAGCCCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1920
QY 1921 ATAGTGGGGGAGATGAGACAGAGTGAATGCTCTCAACAAACGCGTCCGCCACAG 1980
Db 1921 ACAGCTGGGGTCAAAATGATACGGATGCTTCTGCTTAAACACACAGGCCACCGCTGG 1980
QY 1981 GCAACTGGTTCGGCTGATCATGGATGAATAGTACTGGTTCATAGAGCTGGGAGGTC 2040
Db 1981 GCAATTGGTTCGGTGTACTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2040
QY 2041 CCCCCTGTAAATCGGGGGGGTGGTAAACCGACCTTGATCTGCCCCACCGACTGTTCC 2100
Db 2041 CCCCCTGTGTATCGGAGGGTGGGCAACAACACCTTGCTTGCCTCCATGATGCTTCC 2100
QY 2101 GGAAGACCCGAGGCTACTTACAAATGTGGTTCGGGGGCTGGTGGTGGTGGTGGTGGT 2160
Db 2101 GCAACATCCGAAGCCATATCTCTCGGTGGGGTTCGGTCCCTGGATTAACACAGGT 2160
QY 2161 GCTACTGACTACCATACAGCTTTGGCACTACCCCTGCACTCTCAATTTTCCATCT 2220
Db 2161 GATGTGACTTACCCGTTAGGCTTTGGCACTATCTTGTACCAATTAACCAATAT 2220
QY 2221 TTAAGCTTAGGATGTATGGGGGGGTGGAGACACAGGCTCAATGCGGATCAATGGA 2280
Db 2221 TCAAATGTCAGGATGTACGTGGGAGGGTGGAGACACAGGCTGGAGGGCTGCAACTGA 2280
QY 2281 CTCGAGGAGGCTGTAACTGGAGGACAGGATAGGTACAGAACTCAGCCGCTGCTGC 2340
Db 2281 CGGGGGGCAACGCTGTGATGGAAGACAGGACAGGTCCGAGCTCAGCCGCTGCTGC 2340
QY 2341 TGTCTCAACACAGTGGCAGACTGCCCTGTGCTTTTCCACCCCTACCGGCTTTATCCA 2400
Db 2341 TGTCCACCAACAGTGGGAGGCTTCCGTTCTTTTACGACCTGCCAGCCTTGTCCA 2400
QY 2401 CTGGTTGTATCCATCCATCAGAAATCGTGAACGTGCAATACCTGTACGGTGTAGGTT 2460
Db 2401 CCGGCTCATCCACCTCCACCAATTTGGACCTGCACTTGTACGGGGTAGGTT 2460
QY 2461 CAGCGTTTCTCTCTTGCATCAATGGAGTACATCTGTGTTTCTTCTCTCTCTCTCT 2520
Db 2461 CAGCATCGGCTCTGGGCAATTAAGTGGAGTACGTGTTCTCTCTCTCTCTCTCTCT 2520
QY 2521 CAGACCGGCGTGTGCTGCTGTGTGGATGATGCTGTATAGCCCAAGGCTGAGCCCG 2580
Db 2521 CAGACCGGCGGCTGCTCTGCTTGTGGATGATGATTAATCATATCCCAAGCGGAGCGG 2580
QY 2581 CTTAAGAACTTGGTGGTCTCAATGGGCGTTCGTCGGCGGAGCGATGGTATTTCTCT 2640
Db 2581 CTTTGGAGAACCTCGTAATACTCAATGACATCCCTGGCGGAGCGCAAGTCTTGTGT 2640
QY 2641 CTTTCTGTGTCTCTCTGCGCGCTGGTACATTAAGGSCAGGCTGGCTCTCTGGGCGG 2700
Db 2641 CTTTCTGTGTCTCTCTGCTTTGGTGGTATCTGAAGGTAGTGGGTCGCCGAGCGG 2700
QY 2701 CGTATGCTTTTATGGCGTATGGCCGCTGCTCTGCTCTCTTACTTACCGGCTTACCAACAG 2760
Db 2701 TCTACGCCCTTACGGGATGGCCCTCTCTCTGCTCTCTGCTGGCTTGCCTCAGGGG 2760

QY 2761 CTTACGCTTTGGACCGGAGATGGCTGCTCATGTCGGGGGTGGGTTCTTGTAGGCTGG 2820
Db 2761 CATACGCACCTGGACACGAGAGTGGCCGCTGCTGTGGCGGCTTGTCTTGTCCGGTTAA 2820
QY 2821 TATCTTGAACCTTGTGACCACTACTAAGAGTGTCTCTCATAGGCTCATATAGGTGTAC 2880
Db 2821 TGGCGCTGACTCTGTGCGCATATTACAAAGCGCTATACAGTGTGATGGTGGCTTC 2880
QY 2881 AATACTTTATCACCAGAGCCGAGCGCACATGCAAGTGTGGGTCCCCCCCCCTCAACGTT 2940
Db 2881 AGTATTTCTACACAGAGTAGAGCGCACTGCAAGTGTGGTTCGCCCTTCCCCCTCAACGTT 2940
QY 2941 GGGAGGCGCGGATGCGATCATCTCTCTCATGTTGGGTTCTACAGAGTAAATTTTG 3000
Db 2941 GGGGGGGCGGATGCGGCTCATCTTACTCATGTGTAGTACACCCGACCTGTAATTTG 3000
QY 3001 ACATCACCAACTCTCTGCTGCCATCTCGGCCGCTCATGCTCTCAGGCTGGCTAA 3060
Db 3001 ACATCACCAACTCTCTGCGCATCTTTCGACCCCTTTTGAATCTTCAAGCCAGTTGC 3060
QY 3061 CGAGTGTCCCTACTTCTGCGCGCTCAAGGGCTCATTCGTGCAATGATGATTTAGTGGAA 3120
Db 3061 TTAAGTCCCTACTTCTGTCGCGCTCAAGGCTTCTCCGATCTGGCGCTAGCGGGA 3120
QY 3121 AAGTCGCGGGGTCAATATGTCCAAATGTCTTCAAGAGCTGGGCGCTGACAGTA 3180
Db 3121 AGATAGCGGAGGTCAATTAAGTGAATGGCCATCATCAAGTTAGGGCGCTTACTGGCA 3180
QY 3181 CGTAGTTTAAACCATCTTACCCACTGCGGACTGGGCCACCGCGGCTACAGAGCC 3240
Db 3181 CCTATGTATTAACCATCTTACCCCTCTTCGAGACTGGGCGCAACAACGCTCGAGATC 3240
QY 3241 TTGCGGTGGCGGTAGAGCCCTGCTTCTCCGCAATGAGACCAAGATCATACCTGGG 3300
Db 3241 TGGCGTGGCTGTGAACCACTGCTTCTCCGAAATGAGACCAAGCTCATCAGTGGG 3300
QY 3301 GAGCAGACACGCTGCGTGTGGGACATCATCTTGGGTCTACCGCTCTCCGCCCGAGGG 3360
Db 3301 GGGCAGATACCGCGCGCTGCGGTGACATCATCAAGGCTTCCCGCTCTCTCCCGTAGG 3360
QY 3361 GGAGGACATATTTTGGACCGGCTGATGCTCGAAGGGCAAGGGTGGCGACTCTCTTG 3420
Db 3361 GCCAGGATATCTGCTTGGCGACCGCGAGAAATGGTCTCCAGGGTGGAGTTGCTGG 3420
QY 3421 CCCCCTACACGCTTACTCCCAACAAACCGCGGGGTACTTGGTTCGATCATACTAGCC 3480
Db 3421 CGCCCATCAGCGCTACGCCACAGACGAGAGGCTCTCTAGGGTGTATAATCACAGCC 3480
QY 3481 TCACAGCGGGACAGAACACAGCTCGAAGGGAGGTTCAAGTGTCTTACCGCAACAC 3540
Db 3481 TGACTGGCGGGACAAAACCAAGTGGAGGGTCCAGATCGTGTCAACTGTCTACC 3540
QY 3541 AATCTTCTCTGCGACCTGCTCATCAACGCGGTGCTGCTGACTGTCTTACCATGCGGTGCT 3600
Db 3541 AAACCTTCTTGGCACTGCTCATCAATGGGTATGCTGAGCTGTCTACCAAGGGCCGAA 3600
QY 3601 CGAAGACCTTAGCCGCTCCAAAGGTCCTCAACCCAAATGTACACCAATGTAGACTGG 3660
Db 3601 CGAGACCATCGCATCAACCAAGGCTCTGTCATCCAGATGTATACCAATGTAGCAAG 3660
QY 3661 ACCTCGTGGCTGGCAGCGGCCCCCGGGGCGGCTCCATGACACCATGAGCTGTGCA 3720
Db 3661 ACCTTGTGGGCTGGCGGCTCTCTCAAGGTTCCCGCTCATTTGACACCTGTACCTCGGCT 3720
QY 3721 GCTCGGACCTTTACTTGTGTACGACATGCTGATGTCAATTCGGGTGGCGGAGGCG 3780
Db 3721 CTTCCGACCTTTACTTGTGTACGAGGACGCGGATGTCTATTCCTGCGCGGAGGTG 3780
QY 3781 ACAGAGGGGAGTCTATCTCTCCCAAGCGGCTCTCTTACTGTAAGAGGCTCTCTCGGTG 3840
Db 3781 ATAGAGGGGAGTCT 3840
QY 3841 GTCCATTTGCTTGGCCCTTCGGGGGACGCTGTTGGGGCTCTTCCGGGCTGCTGTGTGCAACC 3900

[illegible]

QY 6061 GCCCGGAGAGGGGCTGTGACGTGATGAACCGGCTGTAGCGTTGCGTTTCGCGGGGTA 6120
 Db |||||
 QY 6061 GCCCGGAGAGGGGCGATGATGATGAACCGGCTATAGCGTTTCGCTCCCGGGGGA 6120
 Db |||||
 QY 6121 ACCACGTTCCCTTACGCACTATGTGCTGAGAGCGACGTGACGACGCTGCTACCTAGA 6180
 Db |||||
 QY 6121 ACCATGTTCCCGCACGCACTACGTGCGGAGAGCGATGACGCGCGCGGTCACTGCCA 6180
 Db |||||
 QY 6181 TCCCTCTAGCCTTACCATCACTCAACTGCTGAAGCGGCTCCACCATGATTAATGAGG 6240
 Db |||||
 QY 6181 TACTCAGACGCTCACTGAACCGAGCTCTGAGGCGACTGCATCAGTGGATTAAGCTCGG 6240
 Db |||||
 QY 6241 ACTGCTTAGCGCACTCCGGCTCGTGGCTAAGGGATGTTGGGATGGATATGACCGG 6300
 Db |||||
 QY 6241 AGTGTAACCACTCCATGCTCCGGTTCCTGGCTAAGGACATCTGGACTGGATGCGAGG 6300
 Db |||||
 QY 6301 TGTGTACTGTATCAAGACCTGCGCTCAAGTCCAAAATCTCTGCGCGGTTTACGGGAGTCC 6360
 Db |||||
 QY 6301 TGTGTAGCGCACTTAAAGACCTGGCTGAAGCGCAAGCTCATGCCAACACTGCTGGGATTC 6360
 Db |||||
 QY 6361 CTCTTCTGTCTATGCAACGCGGGTACAGGGAGTCTGCGGGGGGACGGGATCATGCAAA 6420
 Db |||||
 QY 6361 CCTTGTGTTCTTCCAGCGCGGTTATAGGGGGTCTGGCGAGGAGACGGCAATATGCA 6420
 Db |||||
 QY 6421 CCACCTGCCCATCGGAGCAGATCGCCGGAATGTCAAAAACGGTTTCCATGAGGATCG 6480
 Db |||||
 QY 6421 CTGCTGCCACTGTGAGCTGAGATCACTGACATGTCAAAAACGGACGATGAGGATCG 6480
 Db |||||
 QY 6481 TAGGGCTAAGACTGACAGCAACACGTGCGACGGAAGTTCGCCATCAACGATCAACCA 6540
 Db |||||
 QY 6481 TCGGTCTAGGACTGACAGAAATGTGGAGTGGGAGTTTCCCAATTAACGCCCTACACCA 6540
 Db |||||
 QY 6541 CGGACCTTGCAACCTCCCGCGGCCCAATTAATCCAGGGGCTATGCGGGTGGCTG 6600
 Db |||||
 QY 6541 CGGCGCCCTGTATCCCTTCTTGGCCGAACTATTAAGTTGCGGCTGTGGAGGTTGCTG 6600
 Db |||||
 QY 6601 CTGAGGAGTACGTGGAGTTACGCTGTGGGGATTTCCACTACGTGACGGGATGACCA 6660
 Db |||||
 QY 6601 CAGAGGAATACGTGGAGATAAGCGGGTGGGACITCCACTACGTATCGGTTATGACTA 6660
 Db |||||
 QY 6661 CTGACACGTAAAGTGCCTATGCAAGTTCCGGGCCCCGAAATTTCTACGAGGTTGATG 6720
 Db |||||
 QY 6661 CTGACAACTTAAATGCCGTGTCAGATCCCATCGCCCCGAATTTTCAAGAAATTTGAG 6720
 Db |||||
 QY 6721 GAGTGGGTTGACAGGTACGCTCCGGCTGCAAACTCTTCTACGGAGGAGAGTCAAGT 6780
 Db |||||
 QY 6721 GGGTGGCTTACAGAGTTTGGCCCCCTTGAAGCCCTTGTCTGCGGAGGAGGATATCAT 6780
 Db |||||
 QY 6781 TCCAGGTGCGGCTCAACCAATATTTGGTGGGTGCGAGCTCCCATGCGACGCCGAAACCGG 6840
 Db |||||
 QY 6781 TCAGAGTAGACTCCACAGATACCGGTGGGGTGGCAATTAACCTTGGCGAGCCCGAACCGG 6840
 Db |||||
 QY 6841 ACCTAACAGTGTACTTCCATGCTCACCATCCCTCCCAATATACAGCAGACCGCTA 6900
 Db |||||
 QY 6841 ACGTAGCCGTTGACGTCCATGCTCACTGATCCCTCCCATATAACAGCAGAGCGCGCG 6900
 Db |||||
 QY 6901 AGCGTAGGCTGGGTAGAGGTTCTCCCGCTCTTTAGCCAGTCAATAGCTAGCCAGTTGT 6960
 Db |||||
 QY 6901 GGAGAAAGTTGGCGAGAGGTTACCCCTTCTATGCCCAGCTCTCTCGGCTAGCCAGCTGT 6960
 Db |||||
 QY 6961 CTGGGCTTCTTTGAGGCGACATGCACTACCAACATATGCTCCCGGACGCTGACCTCA 7020
 Db |||||
 QY 6961 CCGCTCCATCTCAAGGCAACTGTACCGCAACCATGATCTCCCTGACCGCGAGTCA 7020
 Db |||||
 QY 7021 TCGAGGCCAACCTTCTGTGGCGAGAGATGGCGGAACATCACTCTCGGTGGAGTCA 7080
 Db |||||
 QY 7021 TAGAGGCTAACCTCTGTGAGGCGAGAGATGGCGGAACATCACTCAGAGGTTGAGTCA 7080
 Db |||||
 QY 7081 AGAATAGGTAGTAATTTCTGCACTCTTTGAAACCGCTTCAAGCGGGGGGATGAGAGGG 7140
 Db |||||
 QY 7081 AGAACAAAGTGTGATTTCTGCACTCTTTCATCGCTTGTGGCAGAGGAGATGAGCGGG 7140
 Db |||||

QY 7141 AGATATCCCTCGCGCGGAGATCCCTCGGAAATCCAGGAAGTTCCCTCAGCGTTGCCA 7200
 Db |||||
 QY 7141 AGGTCTCCGTACTCTGAGAAATTTGCGGAAGTCTCGGAGATTCGCCGGGCTTCCCG 7200
 Db |||||
 QY 7201 TATGGGACGCCCCGAGTACAAATCTCTCCACTGCTAGAGTCTTGGAAAGGACCGGACTAG 7260
 Db |||||
 QY 7201 TCTGGCGCGCGCGGACTTACAACCCCCGCTAGTAGAGACGTGGAAAGGCTGACTAG 7260
 Db |||||
 QY 7261 TCCCTCCGCTGTGATACCGGATGCCCAATTCGCCACTACAGAGGCTCTCCAAATACCACTC 7320
 Db |||||
 QY 7261 AACCCACTGTGTCTCATGGCTGCCCTACCACTCCACGGTCCCTCTGTGCGCTCCG 7320
 Db |||||
 QY 7321 CACGAGAAAGAGGACGTTGCTCTGACAGAAATCCAAATGTGTCTTCTGCTTGGCGGAG 7380
 Db |||||
 QY 7321 CTCGMAAAGCGTACGTTGCTCTCAACGATCAACCTATCTACTGCTTGGCGGAG 7380
 Db |||||
 QY 7381 TGGCACTAAGACCTTGGTGTAGTCTCGGATCTCGGCGTTGATAGCGGACGCGGACCG 7440
 Db |||||
 QY 7381 TTTGCCACCAAAAGTTTGGGAGCTCTCAACTTCCGGCATACGGGCGCAATACGACAA 7440
 Db |||||
 QY 7441 CCGTTCTGCACTGCGCTCCGACGAGGTGACAAAGGATCCGACGTTGAGTCTGACTCTCT 7500
 Db |||||
 QY 7441 CATCTCTGAGCGCGCTCTCTGCTGCCCTCCCGACTCCGACGTTGAGTCTCTTCTT 7500
 Db |||||
 QY 7501 CCATGCCCCCTTGAAGGGGAGCGGGGACCCCGATCTCAGCGACGGGCTTTGGTCTA 7560
 Db |||||
 QY 7501 CCATGCCCCCTTGAAGGGGAGCGCTGGGATCCGATCTCAGCGACGGGTCATGGTCA 7560
 Db |||||
 QY 7561 CGGT---GAGTGAGGAGGCTAGTAGGATGTGCTGTCTGCTCAATGCTCAATGCTTATCGTGA 7617
 Db |||||
 QY 7561 CGGTGAGTGTGGGCGGACACGGAAGATGTGCTGTCTCAATGCTTATTCCTGGA 7620
 Db |||||
 QY 7618 CAGGCGCTGTATCAGCCATGCGCTGCGGAGGAAAGTAAAGCTGCCATCAACCCGTGA 7677
 Db |||||
 QY 7621 CAGCGCATCTGTCAACCCGCTGCGTGGAGAAACAAATGCGCCATCAACGACTGA 7680
 Db |||||
 QY 7678 GCACTCTTTGCTGCTCACCACATGCTGTACGCCAACAAATCCCGACGCGAAGCC 7737
 Db |||||
 QY 7681 GCACTCTGTTGTACGCCATCAATGCTGTGTATTTCCACCACTTCAAGCTGTGCTG 7740
 Db |||||
 QY 7738 TCGGCGAGAAAGTCACTTTTGAAGTTGACAGATTCGAAGTCTGATGATCAATACCGGAG 7797
 Db |||||
 QY 7741 AAAGCGAGAAAGTCACTTTGACAGCTCAAGTCTGACAGCCATTTACAGGAG 7800
 Db |||||
 QY 7798 TACTCAAGGAGTGAAGCGAAGCGCTCCACAGTTTAAAGCTAAGCTTCTATCTATAGAG 7857
 Db |||||
 QY 7801 TGCTCAAGGAGTCAAAACGAGCGGCTCAAAAGTGAAGGCTAACCTTCTATCTAGAG 7860
 Db |||||
 QY 7858 AGGCTGCAAGTGTAGCGCCCCACATTCGGGCAATTCGCAATTTGGCTATGCGGCAAG 7917
 Db |||||
 QY 7861 AAGCTTGCAGCTGACGCCCCACATTTAGCCCAATTTCAAGTTTGGCTATGCGGCAAG 7920
 Db |||||
 QY 7918 AGTCCGGAACCTATCCAGAGGCGGTTAAACATCCGCTCCGTTGGGAGGACTTGC 7977
 Db |||||
 QY 7921 AGTCCGTTGCCATGCCAGAAAGCGGCTAGGCCCATCACTCCGTTGGGAAAGACCTTC 7980
 Db |||||
 QY 7978 TGAAGACACTGAAACCAATTTGACACCCATCATGGCAAAAAGTGAAGTTTCTCG 8037
 Db |||||
 QY 7981 TGGAGACACTGTAACACCAATAGACACTACCATCATGGCAGAACGAGGTTTCTCG 8040
 Db |||||
 QY 8038 TCCAACGAGAGAGGAGCGCGAGCAGCTGCGCTTATGCTATTCGACAGCTGGGAG 8097
 Db |||||
 QY 8041 TTAGCGCTGAGAAAGGGGGTCTGAAGCCAGCTCGTCTCATCTGTTTCCCGACCTGGG 8100
 Db |||||
 QY 8098 TTAGCTGATGCGAGAGATGGGCTTTTACGAGTGTCTCCACCTTCTCAGGCGGTA 8157
 Db |||||
 QY 8101 TGGCGGTGTGCGAGAGATGGGCTTGTACGAGTGTGTAGCAAGCTTCCCTTGGCGTGA 8160
 Db |||||
 QY 8158 TGGGCTTCTCATACGATTTCAATCTCCCGCAAGCAGCGGCTGAGTCTCTTGGTGAATA 8217
 Db |||||
 QY 8161 TGGGAGCTTCTACGATTTCCAAATCTACAGAGCAGCGGTTGAATTCCTCGTGCAG 8220
 Db |||||
 QY 8218 CTTGGAATCAAGAAATGCGCTATGCGCTTCTCATATGACACCGCTGTCTTGTGACTCAA 8277
 Db |||||

Db 8221 CGTGAAGTCCAGAGACCCGATGGGTTCTCGTATGATACCGCTGTTTACCTCA 8280
 Qy 8278 CGGTACAGAGAGTACATTCGTGTTGAGAGTCAATTTACCAATGTTGACATGGGCC 8337
 Db 8281 CAGTCACTGAGAGGACATCCGTACGAGAGGAGCAATTTACCAATGTTGACCTGGACC 8340
 Qy 8338 CCGAGGCCAGACAGCCCATAGGTGCTCAGAGAGGCTTTACATCGGGGTCCTCCCTGA 8397
 Db 8341 CCCAAGCCCGGTGGCCATCAAGTCCCTCATGAGAGCTTTATGTTGGGGGCCCTCTTA 8400
 Qy 8398 CTAATCAAAAGGAGCAACTGCGTTATCCCGGTGCGCGCAAGTGGCGTGTGACGA 8457
 Db 8401 CCAATCAAGGGGGAATCGCGTACCGCAGTGGCGCGAGGGCGTACTGACAA 8460
 Qy 8458 CTAGCTGGTAAATACCTCATCTACATTTACAGGCGACTGACGCTGTGAGTCCAA 8517
 Db 8461 CTAGCTGGTAAATACCTCATCTACATCAAGGCGCGGCGAGCTGTGAGCGCGAG 8520
 Qy 8518 AGCTCCAGGACTGCAGATGCTCGTGAACGAGAGCACTTGTGCTTATCTGTGAAGCG 8577
 Db 8521 GGCTCCAGGACTGCACCATGCTCGTGTGGGAGCACTTATGTTCTGTGAAGTG 8580
 Qy 8578 CGGGAACCCAGAGAGTGGCGGCGCTTACAGGCTTCCAGGAGCTATGATAGTATT 8637
 Db 8581 CGGGGGTCCAGGAGGACGCGCGAGCTGTGAGAGCTTCCAGGAGCTATGACAGGTACT 8640
 Qy 8638 CCGCCCCCGGGGATCCGCCCAACAGATAGGCTGAGCTGTATACATCATGTT 8697
 Db 8641 CCGCCCCCGGGGATCCGCCCAACAGATAGGCTGTGAGCTTATACATCATGCT 8700
 Qy 8698 CCTCAATGTGTAGTCCGCGACGATGCTTGGCAAGGATATACCTACCTCCCGGTG 8757
 Db 8701 CCTCCAGTGTAGTCCGCCCAACAGGCTGTGGAAGAGGTCTACTACCTTACCGGTG 8760
 Qy 8758 ACCCCACACCCCGCTTGACGGCTGTGGGAGAGCACTGAGACACACTCCATCAACT 8817
 Db 8761 ACCCTCAACACCCCGCTTGAGAGGCGGCTGGGAGACAGACACACTCCAGTCAAT 8820
 Qy 8818 CTTCGCTAGGCAATATCATCTGATGATGCGGCCACCTATGGCAGAGTATCTGTAGA 8877
 Db 8821 CCTGGCTAGGCAATATCATCTGATGATGCGGCCACCTATGGCAGAGTATCTGTAGA 8880
 Qy 8878 CTCACCTTTTCTCCATCCTTAGTCAAGAGCACTTGAAGAGCCCTGATTTGTCAGA 8937
 Db 8881 CCCATTTCTTAGCGTCTCATAGCCAGGATCAGCTTGAACAGCTCTTAACTGTGAGA 8940
 Qy 8938 TCTAGGGGCTTGTACTTCCATTTAGCCACTTTGACCTTACCTCAGATCATTTGAACGATCC 8997
 Db 8941 TCTAGGAGCTGTACTTCCATAGAACCACTGGATCTACCTCCATCATTTCAAAGACTCC 9000
 Qy 8998 ATGGTCTTAGCGCATTTACATCTCCACAGTTACTCTCCAGGTGAGATCAATAGGTTGGTT 9057
 Db 9001 ATGGCTCAGGCGATTTTCACTCCACAGTTACTCTCCAGGTGAATCAATAGGTTGGCG 9060
 Qy 9058 CATGCTCAGGAACCTTGGGATACCATCTTGCAGAACTCGGAGACATCGGGCCAGAAAGT 9117
 Db 9061 CATGCTCAGGAACCTTGGGATACCATCTTGCAGAACTCGGAGACATCGGGCCAGGAGC 9120
 Qy 9118 TCCGGCTTACTACTTCCAGGGGGGAGGGCGGCACTTGTGGGAGATACCTCTTTTA 9177
 Db 9121 TCCGGCTTACTACTTCCAGAGAGGAGGAGGGCTGCCATATGTGGCAAGTACCTCTTTCA 9180
 Qy 9178 ACTGGGAGTAAAGACCAAGTTTAACTCACTCCATCCCGCGCGCTCCAGCTGGACT 9237
 Db 9181 ACTGGGAGTAAAGACCAAGTTTAACTCACTCCATAGCGGCGCTGGCGCTGGACT 9240
 Qy 9238 TGTCTGGCTGCTTCTGCTGCTTACAGCGGGGAGACATATATCACAGCTGTCTCTGT 9297
 Db 9241 TGTCCGGTTGTTTACAGCTGTGCTACAGCGGGGAGACATTTATCACAGCTGTCTCATG 9300
 Qy 9298 CCGGACCCCGCTGGTTTCGGTTTGGCTTACTTCTTCTTGTAGGGGTAGGGATTTACC 9357

Db 9301 CCGGCCCCCGCTGGTTCTGCTGTTTCTACTCTGCTCGCTGAGGGGTAGGCACTTACC 9360
 Qy 9358 TGCTCCCAACCGATGAAACGGGAGCTTAAACCACTCCAGGCTTAA 9402
 Db 9361 TCCTCCCAACCGATGAAAGTTGGGTAAACACTCCGGCCTCTTA 9405

RESULT 13
 US-09-916-359-1
 ; Sequence 1, Application US/09916359
 ; Patent No. US20020034734A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Veronique Barban
 ; TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
 ; TITLE OF INVENTION: TREATING C HEPATITIS
 ; FILE REFERENCE: PNC97-03A
 ; CURRENT APPLICATION NUMBER: US/09/916,359
 ; PRIOR FILING DATE: 2001-07-26
 ; PRIOR APPLICATION NUMBER: 09/388,874
 ; PRIOR FILING DATE: 1999-09-02
 ; PRIOR APPLICATION NUMBER: 97/02,887
 ; PRIOR FILING DATE: 1997-03-06
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: Fast-Seq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 9379
 ; TYPE: DNA
 ; ORGANISM: Virus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (320) ... (9352)
 US-09-916-359-1

Query Match 64.0%; Score 6144.4; DB 9; Length 9379;
 Best Local Similarity 78.5%; Pred. No. 0;
 Matches 7364; Conservative 0; Mismatches 2011; Indels 3; Gaps 1;

Qy 23 CACTCCACCATGATCACTCCCTGTGAGGAATCTCTTCAACGAGAAAGCGTCTAG 82
 Db 1 CACTCCACCATGATCACTCCCTGTGAGGAATCTCTTCAACGAGAAAGCGTCTAG 60
 Qy 83 CCATGGCGTTAGTATGATGCTGTGAGCTTCCAGAGACCCCTCCCGGAGAGCCATA 142
 Db 61 CCATGGCGTTAGTATGATGCTGTGAGCTTCCAGAGACCCCTCCCGGAGAGCCATA 120
 Qy 143 GTGGTCTGCGGAACCGGTGAGTACACGGAAATTCGAGAGAGCCGCGTCTTCTTGA 202
 Db 121 GTGGTCTGCGGAACCGGTGAGTACACGGAAATTCGAGAGAGCCGCGTCTTCTTGA 180
 Qy 203 TCAACCGCTCAATGCTGAGATTTGGCGGTGCCCCCGGAGAGCTCTAGCCGAGTAGT 262
 Db 181 TCAACCGCTCAATGCTGAGATTTGGCGGTGCCCCCGGAGAGCTCTAGCCGAGTAGT 240
 Qy 263 GTTGGGTGCGAAAGGCGCTTGTGATCTGCTGATAGGGTCTTGGAGTGGCCCGGAG 322
 Db 241 GTTGGGTGCGAAAGGCGCTTGTGATCTGCTGATAGGGTCTTGGAGTGGCCCGGAG 300
 Qy 323 GTCTCGTAGACCGTGCACCATGACGAAATCTTAACCTCAAGAAACCAACCAACGTAA 382
 Db 301 GTCTCGTAGACCGTGCACCATGACGAAATCTTAACCTCAAGAAACCAACCAACGTAA 360
 Qy 383 CACCAACCGCGCCCAACAGGACCTCAAGTTCCCGGGCGGTGGTCAATCGTTGGTGGAGT 442
 Db 361 CACCAACCGCGCCCAACAGGACCTCAAGTTCCCGGGGTGGGGTCAATCGTTGGTGGAGT 420
 Qy 443 TTACTCTTTCGCGGACAGGCGCCAGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 502
 Db 421 TTACTCTTTCGCGGACAGGCGCCAGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 480
 Qy 503 GCGGTGCAACCTGCTGGAAGGGGACCACTATCCCAAGGCTCGCCGACCCGAGGCGAG 562
 Db 481 GCGGTGCAACCTGCTGAGGTAGAGTCAAGCTATCCCAAGGCTCGTGGCGCCGAGGCGAG 540

563	QY	GGCTCGGCTCAGCCGGGTAACCTTGGCCCTCTATGCAATGAGGCGTGGGTGGGC	622
541	Db	GAGCTGGGCTCAGCCGGGTACCTTTGGCCCTCTATGCAATGAGGCGTGGGTGGGC	600
623	QY	AGGATGGCTCTCTCAGCCCGCGCTCCGGCCCTAGTTGGGGCCACAGACCCCGGG	682
601	Db	GGGATGGCTCTCTCAGCCCGCGCTCCGGCCCTAGTTGGGGCCACAGACCCCGGG	660
683	QY	TAGGTCCGTAACTTTGGGTAAGTCAATGAAACCCCTTACATGCGGTTTCGCGATCTCAT	742
661	Db	TAGGTCCGTCAATTTGGGTAAGTCAATGAAACCCCTTACATGCGGTTTCGCGATCTCAT	720
743	QY	GGGGTCAATTCGGCTCTGTCGGCCGCCCTTAGGGGGGCTGCCAGGAGCTTGGSCACAGG	802
721	Db	GGGGTCAATTCGGCTCTGTCGGCCGCCCTTAGGGGGGCTGCCAGGAGCTTGGSCATGG	780
803	QY	TGTCGGGTTCTGGAGAGCGCGTGAACATATGCAACAGGGAACCTTCGCGTTGCTCTTT	862
781	Db	CGTCGGGTTCTGGAAGACGCGTGAACATATGCAACAGGGAACCTTCCTGTTGCTCTTT	840
863	QY	CTCTATCTTCTCTTGGCTCTGTCCTCTTTTGAACATCCAGCTTCGCTTATGAAGT	922
841	Db	CTCTATCTTCTCTTGGCCCTCTCTCTTCTGTTGAGCTTGCCTCTTCGCGCTTACCAAGT	900
923	QY	GCACAACGTTCCGGGATATACCATGTTCAGAAACGACTGCTCCAACTCAAGCAATCTGTA	982
901	Db	GCACAACCTCACGGGCTTTTACCACGTCACCAATGATGCCCTAATCTCGAGTATTGTA	960
983	QY	TGAGGACAGCGGACGTGATCATGCAATACCTCCGGTGGCTGCTCTGTTCAGAGGGTAA	1042
961	Db	CGAGGGCGCCGATGSCCATCTCTCACATCCCGGGTGGCTCCCTTGTCTGAGGGCAA	1020
1043	QY	CAGCTCCGTTGTTGGTTAGCGTCACTCCCAAGCTCGCGCCAGGAATCCCAAGCTCCC	1102
1021	Db	CGCTCGAGGTTTGGTGGCGATGACCCCTACGGTGGCCACAGGATGGCAATCCC	1080
1103	QY	CATAGACAATACGAGCCACGTCGACTTCTCGTTGGACGGCTGCTTTCTGCTCCGC	1162
1081	Db	CGCGACGAGCTTCGACGTCACATCGATCTGCTTTCGGAGCGCCACCTCTGTTCCGC	1140
1163	QY	TATGTACGTGGGGATCTCTCGGATCTATTTTCTCGTCCCTCCAGCTGTCACTTCTC	1222
1141	Db	CCTCTACGTGGGGACCTATGGGGTCTGTTCTTGTGGCCAACTGTTCACTTCTC	1200
1223	QY	GCTTCGCCGCATGACAGCTGACGAGTCAACTGCTCAATCTATCCCGGCATGATC	1282
1201	Db	TCCGAGCGGCACTGGACGACGAAAGTTGCAATGCTCTATCTATCCCGGCATATAAC	1260
1283	QY	AGGTACCGCATGGCTTGGGATATGATGACTGTCTACCTTACACAGCCCTAGTGGT	1342
1261	Db	GGGTACCGCATGGCATGGGATATGATGATGAACTGGTCCCTACGACGGCTGGTAAT	1320
1343	QY	GTCCGAGTTCTTCGGATCCCAAGCTGCTGGACATGGTGGGGGGGCCCACTGGGG	1402
1321	Db	GGCTCAGCTGCTCCGGATCCCAAGCCATCTTGGACATGATCGCTGTCTCACTGGGG	1380
1403	QY	AGTCTTGGCGGCTTGCCTACTATTCATGTGTAGGGAATGGGGCTTAGGTTCTGATGT	1462
1381	Db	AGTCTTGGCGGATACGTAATTTCTCCATGTGTGGGAACTGGCGAAGTCTGTGTGT	1440
1463	QY	GGCGTACTCTTTGCCGGCTTACGGGGAGACCCACAGCGGGGAGGTGGCCGCCCA	1522
1441	Db	GCTGCTGCTATTTGCCGGGCTGACGGGAAACCCAGTCACTCGGGGAGTGGCGGCCA	1500
1523	QY	CACACCTCCGGGTTACGTCCTCTTTTCTCATCTGGGGGCTCTCAGAAAATCCAGTGTGT	1582
1501	Db	CACGTGCTCTGGATTGTTTGAAGCTCTCTCGACACCGGCGCAAGCAGAACTCCAGTGT	1560
1583	QY	GAATACCAACGGGAGCTGGCACATCAACAGGCTGCCCTAAATGCAATCACTCCCTCCA	1642
1561	Db	CAACACCAACGGGAGTTGGCACCTCAATAGCAGGGCCCTGAATGCAATGATAGCTCAA	1620
1643	QY	AACTTGGGTTCTTTGCCCGCTGTTTTTACGCACACAAGTTCAACTGCTCCGGGTGCCGGA	1702

[illegible]

Db 2701 GCCTCTCCTCTGCTCTCTGTTGGCGTTTGCCTCCAGCGGCGCTACGCGCTGGACACGGAGGT 2760
Qy 2783 GGCTGCATCTGTCGGGGGTGGGTCTCTGAGGTCCTGTAATCTCTGACCTTGTCAACATA 2842
Db 2761 GGCCTCTCTGTCGGGTGTTGTTCTCTGTCGGGTGATGCGCTGACTCTGTCAACATA 2820
Qy 2843 CTACAAAGTGTTCCTACTAGGCTCATATGGTGTGTTACAAATCTTTATCACAGAGCCGA 2902
Db 2821 TTACAAGGCTATATCAGCTGGTGTCTGTTGGTGGCTTCAGTATTTCTGACAGAGTGA 2880
Qy 2903 GCGCATACGAGTGGGTCCCTCCCTCAAGTTCGSGGAGCGCGCATGTCAT 2862
Db 2881 AGCGCACTGACGTGTGAGATCCCTCCCTCAAGTCCGAGGGGGCGCGACGCGTCAT 2940
Qy 2963 CCTCCTCACGTGTGGGTTCATCCAGAGTTAAATTTTGAATCAACCAAACTCTCTGCTCGC 3022
Db 2941 CTTACTCATGTGTCTGACACCGACTCTGGTATTTTGACATCACCAAAATGCTGTGCG 3000
Qy 3023 CATACTCGGCGGCTCATGTTGTCCTGAGTCCAGGTGGCATAACGAGTGGCTGCTGTCGCG 3082
Db 3001 CGTCTTCGGACCTCTTTGGATCTCTCAAGCCAGTTCTGCTTAAAGTACCTTACTTTGTGCG 3060
Qy 3083 CGCTCAGGGCTCAATTCGTGCATGTCATGTTAGTCGGAAGTCCCGGGGTCAATATGT 3142
Db 3061 CGTCCAGGGCTCTCTCGGTTCTCGGTTAGCGCGGAAGATGATCGAGGGCCATACGT 3120
Qy 3143 CCAATAGTCTCATGAAGTGGCGGCTGACAGTACGTACCTTTATTAACCACTTAC 3202
Db 3121 GCAANUGTCAATCAATTAAGTTAGGGCGCTTACTTGGCACCTATGTTTATAACCACTCAC 3180
Qy 3203 CCATCTCGGGACATGCGGCCACCGGGCTTACGAGACTTTCGCTGGCGGTAGACCGCT 3262
Db 3181 TCCTCTTCGGGACTCGGGCGCACAAACGGCTTCGAGATCTGGCCCTGCTGTAGAGCCAGT 3240
Qy 3263 CGTCTTCTCGCCATGAGACCAAGTCACTACCTGGGGAGCACACCGCTGCTGTGG 3322
Db 3241 CGTCTTCTCCAAATGAGACCAAGTCACTACCTGGGGAGCATACCGCGCGGTGCGG 3300
Qy 3323 GGCATCATCTTGGGTCTACCCGCTCTCCCGCGGAGGGGGAAGAGATATTTTGGGACC 3382
Db 3301 TGACATCATCAACGGCTTGCTGTTCCGCGCGCAGGGGCGGAGATCTGCTCGGGCC 3360
Qy 3383 GGTGATAGTCTCGAAGGGCAGGGTGGCGACTCTTTCGCGCCATACGCGCTACTCCCA 3442
Db 3361 AGCCGATGGAATGCTCTCCAAAGGGTGGAGGTGCTGCGCGCCATCACGGCTACGCCCA 3420
Qy 3443 ACAACCGGGGCTTACTTGTTCATCATCACTAGCTACAGCTACAGSCCGGGACAGAACCA 3502
Db 3421 GCAGACAAGSGGCTCTAGGGTGCATAATCACAGCTTACTGCGCGGGACAAACCA 3480
Qy 3503 GGTCCGAAGGGAGGTTCAAGTGGTTTCTPACCGCAACACAAATCTTTCTGGCGACTGCAT 3562
Db 3481 AGTGGAGGTTGAGTCCAGATGTTGTCAACTGTCTGCCCAACCTTCTTGGCAACGTGCAT 3540
Qy 3563 CAACGGGTGTGCTGACTGTCTACCATGGCGCTGGCTCGAGACCTTAGCCGCTCCAAA 3622
Db 3541 CAATGGGGTGTGCTGACTGTCTACCGGGGCGGGAAACGAGGACCATCGCTCACCCAA 3600
Qy 3623 AGTTCCAATCAACCAATGTACCAATGTAGACTGGACCTCTGTCGCTGCGAGGCGCC 3682
Db 3601 GGGTCTCTCATCCAGATGTATACCATGTATACAGACCTTGTGGGCTGGCCCGCTCC 3660
Qy 3683 CCGCGGGGCGGCTCCATGACACCATGACAGTGTGGCAGCTCGGACCTTTACTTGGTCAC 3742
Db 3661 GCAAGGTAGCGCTCAATGACACCTGCACTTGGGCTCTCTCGAOCCTTTACTGTTGTCAC 3720
Qy 3743 GAGACATGCTGATGTCATTCGGTGGCGCGGGAGGCGACAGCAGGGGAATCTACTCTC 3802
Db 3721 GAGGACCGCCGATGTCATTCCTCGTGGCGCGGGGTGATAGCAGGGGCAAGCTCTGCTGC 3780
Qy 3803 CCCGAGGSCCGTCTCTTACTGAAAGGCTCTCGGGTGTCTCCATGCTTTGCGCTCGGG 3862
Db 3781 GCCCGGCCCCATTTCTTACTTGAAAGGCTCTCTCGGGGGGTCCGCTGTTGTGCGCGCGGG 3840

Qy 3863 GCACTGTCGTGGGCTCTTCCGGGCTGCTGTGTGCACCCCGGGGGGTCCGCAAGAGCGGTGA 3922
Db 3841 GCAAGCGGTGGGCAATTTAGGGCCGCGGTGTGCACCCGTTGAGTGGCTAAAGCGGTGA 3900
Qy 3923 CTTCAACCGGTGAGTCTATGGAACCTACCATGCGGTCTCCGCTCTTCCAGACAACTC 3982
Db 3901 CTTTATCCCTGTGGAAACCTAGAGAACCACTGAGTCCCGGTGTTCCAGGATAACTC 3960
Qy 3983 AACCCCGGCTGTACCGCAGACATTTCCAAAGTGGACATCTGACGCTCTCTACTGGCAG 4042
Db 3961 CTCTCCACAGTAGTGGCCCGCAGAGCTTCCAGTGGCTACCTCCATGCTCCACAGGAG 4020
Qy 4043 CGCAGAGACCAAAAGTGGCGCTCGTATGACAGCCCAAGGGTACAAGGTGCTCGTCT 4102
Db 4021 CGSCAAAGACCAAAAGTCCCGCTGCATATGACGTACAGGCTATAAGGTGCTAGTACT 4080
Qy 4103 GAAACCGCTCGTTGCGCCACCTTAGGTTTGGGGGTATATGTCCAAAGGACACCGTAT 4162
Db 4081 CAAACCCCTCTGTTGCTGCAACACTGGGCTTTGGTGTCTTACATGTCCAAAGCTCATGGAT 4140
Qy 4163 CGACCTTAAACATCAGAACTGGGGTAAAGACCATTTACACGGGCGGCTCCANTACGTACTC 4222
Db 4141 CGATCTTAAACATCAGACCGGGGTGAGACAAATTTACATGCGCAGGCCCATACGTACTC 4200
Qy 4223 CACTATGGCAAGTTCCTTTCGACAGGTGGCTGTTCTGGGGGGCGCTATGACATCATAT 4282
Db 4201 CACTACGGCAAGTTCCTTTCGACAGCGGCTGCTCGGGGGCGCTTATGACATATAT 4260
Qy 4283 ATGTGATGAGTGCACACTCACTGACTGACTACATCTTGGGGCATCGGCACACTCTGGA 4342
Db 4261 TTGTGAGTGTGCTCACTCCACGGATGCCATCTCTTGGGCATCGGCACACTGCTCTTGA 4320
Qy 4343 CCAAGCGGAGACGCTGGAGCGGCTCGTCTGCTCGCACCGCTACACTCCGGGATC 4402
Db 4321 CCAAGCAGAGACTCGGGGGCGAGACTGTTGTGCTGCGCACCGCCACCTCCGGGCTC 4380
Qy 4403 GGTTCACGTGCCACACCCCAATATCGAGGAAATAGGCTGTCTCCAAACAAATGGAGAGATCCC 4462
Db 4381 CECTCACTGTGCCCATCCCAACATCGAGGAGTTGCTCTGTCTCCACACCGAGAGATCCC 4440
Qy 4463 CTTCTATGGCAAAAGCCATCCCAATGAGGCCATCAAGGGGGGAGGACATCTCATTTCTG 4522
Db 4441 TTTTATCGGCAAGGCTATCCCTCTCGAAGTAAATCAAGGGGGGAGAGCATCTCATCTTCTG 4500
Qy 4523 CCATTCCAAGAAATATGTGACGAGCTCGCGCAAAAGCTCACAGGCTCCGACCTGAACGC 4582
Db 4501 TCATTCAAAGAGAGTGCAGCAATCTCGCCGCAAGCTGTCGCATTGGGCATCATATGC 4560
Qy 4583 TGTAGCATATTAACGGGCGCTTGTATGTCCTGTCATACCGCTTATCGGAGACGTCTGTGT 4642
Db 4561 CGTGGCTACTACCGGCTCTTGCACGTCTCCGTCTATCCCGACCGCGGCGATGTTGTCGT 4620
Qy 4643 CGTGGCAACAGACGCTCTANTGACGGGTTTCACGGGCAATTTTGTGACTCAGTGATCGACTG 4702
Db 4621 CGTGGCAACCGATGCGCTCATGACCGGCTATACCGGCAATTCGACTCGGTGATAGACTG 4680
Qy 4703 CAATACATGTGTCAACCGACAGTCCGACTTCAGCTTGGATCCCACTTCAACCTTCAGACT 4762
Db 4681 CAATACGTGTGTCAACCGACAGTCTGATTTACGCTTGTACCTTACCTTACCATTTAGAC 4740
Qy 4763 GACGACGTCGCCCAAGACCGGTGTGGCTCGCAACCGCGAGGTAGAACTGGCAGGGG 4822
Db 4741 AATCAACGCTCCCGCAGGATGCTCTCTCCCGACTCAACGTCGSGGGAGGACTGGCAGGGG 4800
Qy 4823 TAGAGTGGCATCTACAGTGTGTGACTCCAGAGAACCGCCCTCGGGCATGTTCGATTC 4882
Db 4801 GAAGCCAGGATCTTACAGATTTTGGCACCGGGGAGCGCCCTCGGGCATGTTCGACTC 4860
Qy 4883 TTGGTCTCTGTGAGTGTCTATGACGCGGCTGTGCTTGTGTATGAGCTCACGCCCGCTGA 4942
Db 4861 GTCCGCTCTGTGAGTGTCTATGACGAGGCTGTGCTTGTGTATGAGCTCACGCCCGCGA 4920

4943 QY GACCTCGTGTAGTTGCGGGCTTACCTAAATACACAGGGTTGCCGTTGCGCAGACCA 5002
 4921 Db GACTACAGTTAGCTACGAGCGTACATGAACACCCCGGGCTTCCGTTGTCAGACCA 4980
 5003 QY TCTGAGTTCTGGGAGAGGCTCTTCAAGGCTCACCCACATAGATGCCCATTTCTCTGTC 5062
 4981 Db TCTGAATTTGGGAGGGGCTTTTACAGGCTCACTCATATAGATGCCCACTTTCTATC 5040
 5063 QY CCAGACTAAACAGGACGAGAGACATCTTCTTACCTGGTGGGATATCAAGCTACAGTGTG 5122
 5041 Db CCAGACAAAGCAGAGTGGGAGAACCTTCTTACCTGGTAGGTATCAAGACCACTGGTGTG 5100
 5123 QY CGCCAGGGCTCAAGCTCCACCTCCATCGTGGGACCAATATGGAAGTGTCTATACGGCT 5182
 5101 Db CGCTAGGGCTCAAGCCCTCCCTCCATCGTGGGACAGATGTGGAAGTGTGATTGCGCT 5160
 5183 QY GAAACCTTACCTGACAGGCGCAACACCCCTGCTGTATAGCTAGGAGCGCTGCAAAATGA 5242
 5161 Db CAAGCCACCTCCATGSGGCAACACCCCTGCTATACAGACTGGGCGCTGTTCAAGATGA 5220
 5243 QY GGTCACTCTCACACACCCCATTAATAATACATCATGGCATGATGTCGGCTGAACCTGA 5302
 5221 Db AATACCTCTGACACACCCAGTACCAATATACATCATGATGATGTCGGGACCTGGA 5280
 5303 QY GGTCTGCTACTAGCACCTGGTGTCTGTAGCGGAGTCTTGTGACGCTTTGCGCCATPCTG 5362
 5281 Db GGTCTGCTACAGCACCTGGTGTCTGTGTGGCGGCTCTTGGTGTCTTGGCGCGTATTG 5340
 5363 QY CCTGACGACAGGAGTGTGTCTATTGGGACAGATCATCTTGTCTGGGAGGACCACTGT 5422
 5341 Db CCTGTCAACAGGCTGTGTCTATTGGGACAGGTCGTCTTGTCTGGGAGGCGGCAAT 5400
 5423 QY GGTTCGCGACAGGGAAGTCTCTTACAGGAGTTCGATGAGATGGAAGAGTGTCCCTCA 5482
 5401 Db CATACCTGACAGGGAAGTCTCTTACCGAGTTCGATGAGATGGAAGAGTCTCTGACGA 5460
 5483 QY ACTTCTTATACAGAGGGAATGACGTTCGCGGAGCAATCAAGCAAAAGCGGTGGG 5542
 5461 Db CTTTACCGTACATCGAGCAGGGAATGATGCTCGCGGAGCAGTTCAGCAGGAAGCCCTCGG 5520
 5543 QY GTTGTTCGAAACGGCCACAGCAAGGAGGCTGTCTCCGTTGGTGGAGTCCAAAGTG 5602
 5521 Db CTTCTCTGACACCGCGTCCCGTCAAGGAGGTTATCGCCCTGCTGTCTCAGACCAACTG 5580
 5603 QY GCGAGCCCTTGAGACCTCTGCGGCGAAGCACATGTGGAATTTTCAACAGCGGAATACAGTA 5662
 5581 Db GCAAAACTCGAGACCTCTGCGGCGAAGCATATGTGGAATTTTCAACAGTGGGATACAAAT 5640
 5663 QY CTTAGCAGGCTTATCTCTCTGCTGGAACCCCGGAGATGATCATATGATGGCATTTAC 5722
 5641 Db CTTGGCGGGCTTGTCAACGCTGCTGTGTAACCCCGGCAATGCTTCAATGATGCTTTTAC 5700
 5723 QY AGCTTCTATCACTAGCCGCTTCAACCCCAACACCCCTCTGTTTAAATCTTGGGGG 5782
 5701 Db AGCTGCTGTACAGGCCCATTAACCATAGCCAAACCTCTCTTCAACATATGGGGG 5760
 5783 QY ATGGGTGCTTCCCAACTCTCTCTCTCCAGCGTGTCTGAGCTTTCTGCGGCGCCGSCAT 5842
 5761 Db GTGGGTGCTTCCAGCTCTCCGCGCCCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5820
 5843 QY CGCCGAGCGGCTGTTGGAGCATAGGCTTGGGAGGCTGCTGCTGACATCTGGGGG 5902
 5821 Db AGTGGGCGCCCATCGGAGTGTGAGCTGGGAGGCTCTCATAGACATCTCTTGCAGG 5880
 5903 QY CTATGGGCGAGGGTGTAGCGGCGCACTCTGCGCTTTTAAAGTCAATGAGCGCGAGGTGCC 5962
 5881 Db GTATGGCGGGCGGTGGGGAGTCTTGTGGCAATTCAGATCATGAGCGGTGAGTCTCC 5940
 5963 QY CTCACCGGAGGACCTGTCTACTCTCTCTCTGCGCATCTCTCTCTGCTGCTGCTGCTGCT 6022
 5941 Db CTCACCGGAGGACCTGTCTACTCTCTCTCTCTGCGCATCTCTCTCTGCGCGGAGCCCTCGTAGT 6000
 6023 QY CGGGGTGCTGTGCGGACCAATACTGCGTCCGACGTCGGGCGCGGAGAGGGGCTGTGCA 6082

6001 Db CGCGCTGTCTGTGACCAATACTCGCGCGCACGTTTGGCCCGGCGAGGGGCGAGTCA 6060
 6083 QY GTGATGAACCGGCTGATAGCGTTTCGCTTCGCGGGTAAACACGCTCTCCCTACGCACTA 6142
 6061 Db GTGATGAACCGGCTGATAGCGTTTCGCTTCGCGGGGAAACATGTTTCCCGCACGCACTA 6120
 6143 QY TGTGCTGAGAGCGACGCTGACGACGCTGTCACTCAGATCTCTCTCTAGCTTCAAGACCTG 6202
 6121 Db CGTCCGAGAGCGATGAGCTGCGCGCTCACTGCCATATCTCAGACGCTCACTGTAAAC 6180
 6203 QY TCACTGCTGAAGCGGCTTCACCACTGATTAATGAGGACTGCTTACGCGCATGCTCGG 6262
 6181 Db CCAGCTCTGAGGCGACTTCACCACTGATTAAGCTCGAGGTGACCATCTCCATGCTCCGG 6240
 6263 QY CTCGTGGCTTAAGGATGTTTGGGATGATGATGATGATGATGATGATGATGATGATGATG 6322
 6241 Db TTCTGCTGATAGGACATCTGGGACTGATGATGATGATGATGATGATGATGATGATGATG 6300
 6323 QY SCTCCAGTCCAACTCTCTGCGCGGTTACCGGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 6382
 6301 Db GCTAAAGCTAAGCTCATGCGCACAGCTGCTGGGATCCCTTTGTCTCTCTCTCTCTCTCTCT 6360
 6383 QY GTACAGGAGTCTGCGCGGGGAGCGCATCATGCAACCACTCTGCCATTCGCGAGCACA 6442
 6361 Db GTATTAAGGGGTCTGCGGAGTGGACGCGCATCATGCAACCACTCTGCCATTCGCGAGCTGA 6420
 6443 QY GATCCCGGACATGTCAAAACCGGTTCCATGAGGATCGTAGGGCTTAGAACTCTGACGAA 6502
 6421 Db GATCACTGGACATGTCAAAACCGGACGATGAGGATCGTGGTCTTAGGACCTGACGAA 6480
 6503 QY CAGCTGGCACGGAACGTTCCCTCCATCAACGATCAACGAGGACCTTGCACACCTCCGCC 6562
 6481 Db CATGTGGAGTGGGACCTTCCCTCCATTAATGCTTACACCGGCGGCTCTACCGCCCTTCC 6540
 6563 QY GCGGCCCAACTTTCAGGCGCTATGCGGGTGTCTGTAGGAGTGTAGTGTAGGAGTAC 6622
 6541 Db TCGCGCGCAACTACAGCTTGGCGCTATGAGGGTGTCTGTAGGAGGATGTGTAGGAGTAAG 6600
 6623 QY GCGTGTGGGGATTTCCACTGATGACGGGCAATGACCACTGACAAAGTAAAGTCCCAATG 6682
 6601 Db CAGGTGGGGGCTTCCACTGATGACGGGTAATGACTACTGACATCTCAATGCCCCG 6660
 6683 QY CAGGTTCCGGCCCCCGAATTTTCAOGGAGTGTGATGAGTGTGCGGTGCGACAGTACGC 6742
 6661 Db CAGGTGCCCATCGCCCGAATTTTCAAGATTTGACGCGGCTGCGCTACATAGTTTGC 6720
 6743 QY TCGGCGTCAAAACCTTCTTACGGGAGACGCTCACTGTTCCAGGTGCGGCTCAACCAATA 6802
 6721 Db GCGCCCTGCAAGCCCTTCTGCGGAGGAGGATCAATTCAGAGTAGGACTCCACGATA 6780
 6803 QY CTTGTGCGGTGCGAGCTCCCATGCGAGCCCGAACCGGACGTAACAGTGTCTTACTTCCAT 6862
 6781 Db CCGGTGAGGTGCGAATTTACTTGGAGCCGAAACCGGACGTTGGCGCTGTGACGCTCAT 6840
 6863 QY GCTCAGCATTCCTCCCATTTACAGCAGACGCGTATAGCTAGCTGCGGTAGAGGCTC 6922
 6841 Db GCTCACTGATTCCTCCCATTTAACAGAGGCGCGCGCGGAGGTTGGCGGGGATC 6900
 6923 QY TCCCGCTCTTTAGCCAGCTCATGAGTGTGCTGCGGCTTCTTGTGAAGCGGAC 6982
 6901 Db ACCCGCTCTGTGGCCAGCTCTCTCGGTAGCAGCTATTCGCTCCATCTCTCAGGCAAC 6960
 6983 QY ATGCACTACCCCAATGATCTCCCGGACGCTGACCTCATCTGAGGCGCAACCTCTTGTGGG 7042
 6961 Db TTGACCGGTAACTGACTTCCCTGATGCTGAGCTCAUAGAGGCCAACCTCTCTATGAG 7020
 7043 QY GCGGAGATGGCGGGAACATCACTGCGGTGAGTCAAGAGATTAAGTATGATTTCTGGA 7102
 7021 Db GCGAGGATGGCGGGAACATCACTGAGGTTGAGTCAAGAAACAAAGTGTGATTTCTGGA 7080
 7103 QY CTCTTTCGAACCGCTTCAACGCGGAGGGATGAGAGGAGATATCCGTCCGCGCGGAGAT 7162

Db 7081 CTCCTTCGATCCGCTTGTGGCGAGGAGGAGCGGAGATCTCCGTACCCGCGAGAAT 7140
Qy 7163 CCTCGAAAATCCAGGAAGTTCCTCTCAGCGTGGCCATATGGGACGCCCGGACTACAA 7222
Db 7141 CCTCGGAAAGTCTGGAGATTCGCCAGGCCCTGCCGTTTGGGGCGCGCGACTATAA 7200
Qy 7223 TCCTCCACTGCTAGAGTCTCTGGAAGGACCGGACTACGTCCTCTCGGTGGTACAGGATG 7282
Db 7201 CCCCCCGTAGTGGAGAGTGGAAAAAGCCGACTACGAACCACTGTGTGTCATGGCTG 7260
Qy 7283 CCCATGCACTACCAAGGCTCTCCAAATACCACTCCACGGAGAAAGAGACGCTTGT 7342
Db 7261 TCCGCTTCCACTCCAAAGTCCCTCTCTGCGCTCCGCTCGGAGAGAGCGAGCGTGT 7320
Qy 7343 CTTGACAGATCCAAATGTCTCTGCTTGGCGTGGCGAGCTCGCACTAAGACTTCGGTAG 7402
Db 7321 CTTCACTGAATCAACCTATCTACTGCTTGGCGAGCTCGCCACCAAGATTTGGCAG 7380
Qy 7403 CTCGGATCGTGGCGCTGTGATAGCGGACCGCGACCGCCCTCTCTGACCTGGCTCCGA 7462
Db 7381 CTCCTCAACTTCGGGCAITACGGGCGACATACGACAAACATCTCTGAGCCCGCCCTTC 7440
Qy 7463 CGACGCTGACAAAGATCGAGCTTGAAGTCTGCTCTCCATGCCCCCTTGAAGGGA 7522
Db 7441 TGGCTGCCCCCGGACTCCGAGCTCGAGCTGAGTCTATTCTCCATGCCCCCTTGGAGGGA 7500
Qy 7523 GCGGGGACCCCGATCTCAGCGACGGGTCTGTGTCTACCGTGAAGGAGGCTA--G 7579
Db 7501 GCTTGGGATCGGATCTTAGCGAGGGTCACTGTCAACGCTCAGTGTAGGCGCAACGC 7560
Qy 7580 TGAGATGTGCTGCTCTCAATGCTCTAATGCTGTAAGTGGACAGCGCCCTGATCAGCCAT 7639
Db 7561 GGAGATGTGCTGCTCTCAATGCTCTAATGCTTACTTCTTGGACAGCGGCACTCGTCA 7620
Qy 7640 CGCTCGGAGAGAGTGAAGTCCCATCAACCGCTTGAAGACTCTTGTCTGGCTACCA 7699
Db 7621 GCGCGGAGAGAGAGAGTCCCATCAATGCTAAGCACTGCTGCTACGTCACCA 7680
Qy 7700 CAACATGTCTACGCCACAAATCCCGCAGCGCAAGCTCCGGCAGAGAGAGTCACTT 7759
Db 7681 CAATTTGTGTATTCACCACTCAGCAGTGTCTTGGCAAGGCGAGAGAGTCAAT 7740
Qy 7760 TGACAGATGAAGTCTGGATGATCAATACGGGACCTACTCAAGGAGATGAAGCGAA 7819
Db 7741 TGACAGATGAAGTCTGGACAGCCATTAACAGGACCTACTCAAGGAGGTAAAGCAGC 7800
Qy 7820 GCGTCCACAGTGAAGTGAAGTCTCTATCTATAGAGAGGCTCCAGCTCAGCCGCC 7879
Db 7801 GCGTCAAAAGTGAAGGCTAATCTGCTATCGTAGAGAGCTTGCAGCTGAGCCGCC 7860
Qy 7880 ACATTCGCCCAATCCAAATTTGGCTATGGGCAAGGACGCTCCGGAACCTATCCAGCAG 7939
Db 7861 ACATTCGCCCAATCCAAATTTGGTATGGGCAAGGACGCTCCGTTCCATGCCAGAA 7920
Qy 7940 GCGGTGAACACATCCGCTCGTGTGGAGGACTTGTCTGGAAGCATCTGAAAACCAAT 7999
Db 7921 GCGGTGAACACATCAATCCGCTGTGGAAGACCTCTCTGGAAGCAATGTAACACCAAT 7980
Qy 8000 TGACACCACTATGAGCAAAAGTGAAGTCTTCTGCTCCAAACAGAGAGGAGGCG 8059
Db 7981 AGACATPACCATCTAGCTAAGACAGGTTTCTGCTTACGCTGAGAGGGGGTGC 8040
Qy 8060 CAAGCAGCTCGCTTATCGTATTCAGACCTGGGAGTCTGTATGCGAGAGATGCC 8119
Db 8041 TAAGCAGCTGCTCTCATCGTGTTCCTGATCTGGGCGTGGCGTGTGGAAGATGCG 8100
Qy 8120 CTTTACAGCTGCTCTCAACCTTCTCAGGCTGATGGGCTCCTCATACGATTTCA 8179
Db 8101 TTTGTACAGCTGCTTCAAAAGCTCCCTTGGCGCTGATGGGAAGCTCTACGATTTCCA 8160
Qy 8180 ATACTCCCCAAGACGGGTCGATCTCTGCTGATATCTGGAATACCTGGAAATCAAGATGCC 8239
Db 8161 ATACTCAACAGACGGGTTGAATTTCTCTGCAAGCTGTGAAGTCCAGAAACCCC 8220

Qy 8240 TATGGGCTTCTCATATGACACCCGCTCTTTTGAATCAACGTCATGAGAGTGACATTCG 8299
Db 8221 AATGGGTTCTCGTATGATACCCGCTGCTTTGATCTCAGTCTACTGAGAGCGACATCCG 8280
Qy 8300 TGTGAGAGTCAATTTACCAATGTTGTGACTTTGGCCCCCGAGGCCAGACAGGCAATAAG 8359
Db 8281 TACGGAGGAGCAATCTACCAATGTTGTGACCTCGACCCCAAGCCCGCGTGGCACTAA 8340
Qy 8360 GTCCCTCACAGAGCGGCTTTACATCGGGGGTCCCTGACTTAACCTCAAAAGGAGAACTG 8419
Db 8341 GTCCCTCACAGAGAGGCTTTATGTTGGGGGCTCTTACCAATTCAGGGGGGAGAACTG 8400
Qy 8420 CGGTATTCGCGGTCGCGCAAGTGGCTGTGAGACTAGCTAGCTGCGGTATATACCTCAC 8479
Db 8401 CGGTATTCGAGGTCGCGCGAGCGGTACTGACAACTAGCTGTGTGTAACACCTCAC 8460
Qy 8480 ATGTTACTTGAAGCCACTGTCAGGCTGTGAGCTGTGAAAGCTCCAGACTGACGATGCT 8539
Db 8461 TTGCTACATCAAGCCCGGCGAGCTGTGAGCGCGAGGGTCCAGGACTGCACATGCT 8520
Qy 8540 CGTCAACGGAGACGACTTGTGCTTATCTGTGAAGCGCGGAACCCAGAGAGATCGGC 8599
Db 8521 CGTGTGGCGACGACTTAGTGTATTCTGTGAAGCGCGGGTCCAGAGAGCGCGC 8580
Qy 8600 GGCCTTACGAGCCCTTCAGGAGGCTATGACTAGGTATTTCCGCCCCCCCCCGGATTCGCC 8659
Db 8581 GAGCTTGAAGGCTTCAGCGAGGCTATGACAGGACTCTCCGCCCCCTTGGGACCCCC 8640
Qy 8660 CCACACAGAAATCAGCTTGAAGCTGATGATCAATGTTTCTTCAATGTGTGTCAGTGGCA 8719
Db 8641 ACAACAGAAATACGACTTGAAGCTATGATCATATGCTCTCCAACTGTGTGTCAGTGGCA 8700
Qy 8720 CGATCATCTGGCAAAAGGCTATGACTACCTCAACGCTGACCCACACCCCTTGGCAG 8779
Db 8701 CGAGCGCTGGAAAGAGGCTTACTACCTCAACGCTGACCCCTCAACCCCTTGGCAG 8760
Qy 8780 GGTCTGCTGGAGACAGCTACACACTCCAATCAACTCTTGGCTAGGCAATATCATCAT 8839
Db 8761 AGCTGCTGGAGACAGACACACTCCAAGTCAATCTCTGGCTAGGCAATATCATCAT 8820
Qy 8840 GTATGCGGCCACCTATGGGCAAGGATGATTTGATGACTCACTTTTCTTCTTCTTCT 8899
Db 8821 GTTTGCCCCACACTGTGGCGAGGATGATCTGATGACCACTTTTCTTGTAGCTCTTAT 8880
Qy 8900 AGCTCAAGAGCAACTTCAAAAGGCTGATGTCAGATCTAGCGGCTTGTACTCAT 8959
Db 8881 AGCCAGGACAGCTTGAAGGCGCTCGATTTGGAGATCTACGGGCTTGTACTCAT 8940
Qy 8960 TGAGCCACTTGACCTACTCAGATCATTTGAACGACTCCATGCTTGTAGCCATTTTACT 9019
Db 8941 AGAACCTTGTACTCTCTCAATCATTTCAAGACTCCATGGCTCAGGCACTTTTCACT 9000
Qy 9020 CCACAGTACTCTCAGGAGATCAATAGGTGGCTTCTATGCTCAGGAAACTTGGGT 9079
Db 9001 CCACAGTACTCTCAGGAGAAATTAATAGGTGGCGATGCTCAGAAACTTGGGT 9060
Qy 9080 ACCACCTTGGAGACTTGGAGACTCGSGCCAGAGTGTCCGCTAAGCTTGTGCCCCA 9139
Db 9061 ACCGCTTGGAGCTTGGAGACTCCGGGCCCGAGGCTCCGCTTAGGCTTCTGGCCAG 9120
Qy 9140 GGGGGGAGGCGCCCTTGTGGCAGATACCTCTTTTAACTGGGSCAGTAAGGACCAAGCT 9199
Db 9121 AGGAGGCGGGCTGCCATATGTGCAAGTACTCTTCACTGGGSCAGTAAGAACCAAGCT 9180
Qy 9200 TAAATCACTCCAAATCCCGCGCGCTCCAGCTGGAATTTGTCTGGCTGGTTCGCTGG 9259
Db 9181 CAAACTCACTCCAAATAGCGCGCTGGCCAGCTGGAATTTGTCTGGCTGGTTCAGCGCTG 9240
Qy 9260 TTACAGCGGGGAGACTATATCAAGCTGTCTCGTGGCCGACCCCGCTGCTTCCGT 9319
Db 9241 CTACAGCGGGGAGACATTTATCAAGCTGTCTCATGCGCGGCCCGCTGATCTGCT 9300

QY	9320	GTGCGTACTCTACTTCTCTAGGGGTAGGGAATTACTGTCCTCCCAACCGATGAACGGG	9373
Db	9301	TTCGCTACTCTCTGCTTCCTCAGGGGTAGGCATACCTCTCTCCCAACCGATGAAGTT	9360
QY	9380	GAGCTAACCACTCCAGCG	9397
Db	9361	GGGTAAACACTCGGCC	9378
RESULT 14			
US-09-238-076-19			
; Sequence 19, Application US/09238076			
; Patent No. US20020102340A1			
; GENERAL INFORMATION:			
; APPLICANT: RICE, CHARLES et al.			
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C			
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF			
; NUMBER OF SEQUENCES: 21			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.			
; STREET: 7733 FORSYTH BLVD., SUITE 1400			
; CITY: ST. LOUIS			
; STATE: MO			
; COUNTRY: USA			
; ZIP: 63105			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/238,076			
; FILING DATE:			
; CLASSIFICATION:			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 09/034,756			
; FILING DATE:			
; ATTORNEY/AGENT INFORMATION:			
; NAME: HOLLAND, DONALD R.			
; REGISTRATION NUMBER: 35,197			
; REFERENCE/DOCKET NUMBER: 6029-4831			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 314-727-5186			
; TELEFAX: 314-727-6092			
; INFORMATION FOR SEQ ID NO: 19:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 9416 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: linear			
; MOLECULE TYPE: cDNA			
; HYPOTHETICAL: NO			
US-09-238-076-19			
Query Match 63.9%; Score 6135,6; DB 9; Length 9416;			
Best Local Similarity 78.3%; Pred. No. 0;			
Matches 7372; Conservative 0; Mismatches 2039; Indels 3; Gaps 1;			
QY	1	GCACGCCCCCTGATGGGGCGACACTCCACCAATCACTCCCTCTGTGAGGAACTACTG	60
Db	1	GCACGCCCCCTGATGGGGCGACACTCCACCAATCACTCCCTCTGTGAGGAACTACTG	60
QY	61	TCCTCAGCAGAAAGCGTCTAGCCATGGCGCTTAGTAGTGTCTGCGAGCTCCAGGAC	120
Db	61	TCCTCAGCAGAAAGCGTCTAGCCATGGCGCTTAGTAGTGTCTGCGAGCTCCAGGAC	120
QY	121	CCCCCTCCCGGAGAGCCATAGTGGTCTGCGAAACCGGTGAGTACACCGGAATTCGCCAG	180
Db	121	CCCCCTCCCGGAGAGCCATAGTGGTCTGCGAAACCGGTGAGTACACCGGAATTCGCCAG	180
QY	181	GACGACCGGGTCTCTTCTTGATCAACCGCTCAATGCTTGGAGATTGGGGGTGCCCC	240
Db	181	GACGACCGGGTCTCTTCTTGATCAACCGCTCAATGCTTGGAGATTGGGGGTGCCCC	240

Dd 3481 TGACTGGCGGAGCAAAACCAAGTGGAGGGTGAGGTCCAGATCGGTCAACTGCTACCC 3540
Qy 3541 AATCTTTCTGCGACCTCATCAACCGCGTGTGCTGGAGTCTACCAATGCGGTGGCT 3600
Dd 3541 AAACCTTCTGCAAGTGCATCAATGGGTATGCTGGACTCTTACCAAGGGGCGGAA 3600
Qy 3601 CGAAGACCTTAGCGGTTCAAAAGTCCAAATCACCCAAATGACCAATGTAGACCTGG 3660
Dd 3601 CGAGGACCATCGCATCACCAAGGGTCTGTTCATCCAGATGATACCAATGTGGACCAAG 3660
Qy 3661 ACCTCGTGGTGGAGGCGCCCGCGGCGGTCCATGACACCATGACGTGGCA 3720
Dd 3661 ACCTTGTGGGTGGCGCGCTCTCAAGGTTCGGGTCCATGACACCTCGGCT 3720
Qy 3721 GTCGACCTTTACTGTGTCACGAGACATGCTGATGTCATTCGGTGGCGCGGAGGCG 3780
Dd 3721 CCTCGACCTTTACCTGGTTACGAGGACGCGACGTCAATCCCGTGGCGCGGAGGTG 3780
Qy 3781 ACAGCAGGGAGTCTACTCTCCCGACGCGCGTCTCTACTGAAAGGCTCCTCGGTTG 3840
Dd 3781 ATAGCAGGGGTAGCTGTCTTCCCGCGCGCCATTTCCCTACCTAAAGGCTCCTCGGCG 3840
Qy 3841 GTCCATTGCTTTCCTTGGGCGACGCTGCTGGGCGTCTTCGGGCTGCTGTGTGACCC 3900
Dd 3841 GTCCGCTGTGTGCGCGCGGACGCGCTGGGCTTATCAGGGCGCGGTGTGACCC 3900
Qy 3901 GGGGGTCCGAGCGGTGGACTTCATACCCGTTGAGTCTATGAAACTACCATGCGGT 3960
Dd 3901 GTGGAGTACCAAGCGGTGGACTTTATCCCTGTGGAAACCTAGAGACAACCATGAGAT 3960
Qy 3961 CTCGGTCTTCACAGACAACCTCAACCCCGCGCTGTACCGGACATTCNAHTGGAC 4020
Dd 3961 CCCCCTGTTTACGGACAACTCTCTCCACCAAGCACTGCCCCAGAGCTTCAGGTGGGCC 4020
Qy 4021 ATCTCAGCTCTACTTGGGAGCGGCAAGAGACCAAGTGTCCGCTGGGTATGAGGCC 4080
Dd 4021 ACCTGATGCTCCACCGGAGTGTAGAGACCAAGTCCCGGCTCGTACGAGGCC 4080
Qy 4081 AAGGTACAAGTGTCTGTCTGAACCCGTCCGTTGCCGCCACCTTAGCGTTGGGGCT 4140
Dd 4081 AGGGTACAAGGTGTGGTGTCAACCCCTCTGTGTGTCGAAGCTTGGGTGCTT 4140
Qy 4141 ATATGTCAAGGCACACGCTATCGACCTCACTCAGACTGGGTAGGACCATACCA 4200
Dd 4141 ACATGTCAAGGCCATGGGTGCTGATCTTAATATCAGACCGGGGTGAGAACATACCA 4200
Qy 4201 CGGCGCTCCATTAAGTACTCCACTATGCAAGTTCCTTTGCCGACGCGTGGTCTG 4260
Dd 4201 CTGGCAGCCCATACGTAATCCACTACGCAAGTTCCTTGGCCGACGCGGGTCTCAG 4260
Qy 4261 GGGGGCTATGACATCAATATATGATGATGATGATGATGATGATGATGATGATGAT 4320
Dd 4261 GAGCGCTTATGACATAATAATTTGTGACGATGCTCACTCCACGATGCCACATCCATCT 4320
Qy 4321 TGGGATCGGCACATGCTGACCAAGCGGAGAGCGCTGGAGCGGCTCTGCTGCTG 4380
Dd 4321 TGGGATCGGCACATGCTGACCAAGCGGAGAGCTGGGGGCGAGATGTTGTGCTG 4380
Qy 4381 CCACTGTACACTTCGGGTACGTTACCGTCCACACCCCAATATCGAGGAATAGGCC 4440
Dd 4381 CCACTGTACACTTCGGGTACGTTACCGTCCACACCCCAATATCGAGGAATAGGCC 4440
Qy 4441 TGTCCNACATGGAGAGTCCCTTCTATGGGAAGCCATCCCATTTAGGCCATCAAG 4500
Dd 4441 TGTCCNACATGGAGAGTCCCTTCTATGGGAAGCCATCCCATTTAGGCCATCAAG 4500
Qy 4501 GGGGAGGACATCTCATTTCTGCAATCCAAAGAGAAATGACAGAGTCCCGCAAGC 4560
Dd 4501 GGGGAGGACATCTCATTTCTGCAATCCAAAGAGAAATGACAGAGTCCCGCAAGC 4560
Qy 4561 TGAAGGCTCGGATGAAGCTGTAGCATATTACCGGGGCTTGAATGTCCGTGCTATC 4620
Dd 4561 TGGTGGATGGGCATCAATGCGGTGGCTTACTACCGCGGACTTGAAGTGTCTGTATCC 4620

Qy 4621 CGCCTATCGGAGACGTCTGTTGTCGTGCAACAGACGCTCTAATGACGGGTTTCAACGGG 4680
Dd 4621 CGACACGGCGATGTTGCTGCTGTCGACCGATGCTCTCATGACTGGCTTTACCGGCG 4680
Qy 4681 ATTTGACTCAGTGAATGACTGCAATACATGTGTACCCAGACAGTGGACTTCAGCTTGG 4740
Dd 4681 ACTTCGACTCTGTGATAGACTGCAACAGTGTGTCTCATGACAGTGCATTTAGACCTTG 4740
Qy 4741 ATCCACCTTACCATTTAGACGAGCGCGGTGCCCCCAAGACGCGGTGTGCGGTCCGAAC 4800
Dd 4741 ACCCTACTTTTACCATTTAGACCAACACGCTCCCGCGGATGCTGTCTCCAGGACTCAGC 4800
Qy 4801 GCGGAGGTAGAACTCTGGCAGGGTAGGAGTGCATCTACAGATTTGTGACTTCCAGGAAAC 4860
Dd 4801 GCGGGGACGAGACTGGCAGGGGGAAGCAGGCAFTTACAGATTTGTGGACCGGGGAGC 4860
Qy 4861 GGCCTCGGGCATGTTCAATTTCTTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Dd 4861 GGCCTCGGGCATGTTGTGACTGTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Qy 4921 GGTATGAGCTCACCGCGCTGAGACCTCGGTTAGGTTCGGGCTTACTAAAATACACAG 4980
Dd 4921 GGTATGAGCTCATCCCGCGGAGACTACAGTTAGGCTACGAGCGTACATGAACACCCCGG 4980
Qy 4981 GGTTCGCCGTCTGCAGACCATCTGGAGTTCTGGGAGAGCGTCTTCAAGGCGCTCACCC 5040
Dd 4981 GGCTTCGCGTGTGCAGAGACCAATTTGAAATTTGGGAGGGGCTTTTACGGGCTCACCC 5040
Qy 5041 ACATAGATGCCACTTCTGTGCCAGACTAAACAGGAGGAGAGACACTTTTCTTACTCTGG 5100
Dd 5041 ATATAGATGCCACTTCTATCCAGACAAAGCAGAGTGGGGAGAACTTTCTTACTCTGG 5100
Qy 5101 TGGCATATCAGCTACAGTGTGCGCCAGGCTCAGCTCCACCTCCTCATCTGGGACCAAA 5160
Dd 5101 TAGGTAACCAAGCACCGTGTGGCTTAGGCTCAAGCCCTTCCCCATCTGGGACCAAGA 5160
Qy 5161 TGTGGAAGTGTCTCATAGGCTGAAACCTTACCTGACAGCGGCCAACACCCCTTGTGTATA 5220
Dd 5161 TGTGGAAGTGTCTGATCGGCTTAAACCCACCTCTCATGGGCCAACACCCCTTGTGTATA 5220
Qy 5221 GGTAGAGCGGTCCAAAATGAGGTCTATCTTACACACCCCATAACTAAATATCATATGG 5280
Dd 5221 GACTGGCGGTGTGTAGAAATGAAAGTACCCCTGACGACCCCAATCAGCAATATCATATGA 5280
Qy 5281 CATGCTATGCGGCTGACTGAGTGTGCTACTAGCACTGCTGCTGCTGCTGCTGCTGCTGCT 5340
Dd 5281 CATGCTATGCGGCGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5340
Qy 5341 TTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5400
Dd 5341 TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5400
Qy 5401 TCTGTTCGGGAAGCAGCTGTGTGTCGCGACAGGAGTCTTCTTACCGAGATTCGATG 5460
Dd 5401 TCTGTTCGGGAAGCAGCTGTGTGTCGCGACAGGAGTCTTCTTACCGAGATTCGATG 5460
Qy 5461 AGATGGAAGAGTGTGCTTCAACTTCTTACATCGAGCAGGGAATGCACTGCTGCTGCTGCT 5520
Dd 5461 AGATGGAAGAGTGTCTTACAGCATTTACGCTACATCGAGCAAGGATGATGCTCGCTGCTGCT 5520
Qy 5521 AATTCAGCAAAAGGCGCTCGGCTTGTGCAACGGGCCAACCAAGCAAGCGGCTGCTGCTGCT 5580
Dd 5521 AGTTCAGCAAAAGGCGCTCGGCTTGTGCAACGGGCCAACCAAGCAAGCGGCTGCTGCTGCT 5580
Qy 5581 CTCCGCTGCTGAGTCCAGTGGGAGCCCTTGAAGCTTCTGAGGAGAGTCTTCTGAGGAGTTCGATG 5640
Dd 5581 CCCTGCTGCTGAGTCCAGTGGGAGCCCTTGAAGCTTCTGAGGAGAGTCTTCTGAGGAGTTCGATG 5640
Qy 5641 ATTTTCATCAGCGAAATACAGTACCTAGCAGCTTATCCACTCTGCTGCTGCTGCTGCTGCTGCT 5700
Dd 5641 ATTTTCATCAGTGGATACAAATATTTGGCGGCGCTTCAAGCGCTGCTGCTGCTGCTGCTGCTGCT 5700

QY 5701 TAGCATATGATGGCATTTAAGCTTTATACATAGCCCGCTACACACCCAAAACACC 5760
Db 5701 TTGCTTCATGATGGCTTTTAAAGCTGGCGTCAACAGCCCACTAAACCACTGGCCAAAACC 5760
QY 5761 TCTGTTTAACTCTTGGGGGAGTGGTGGCTGCCCAATCGCTCTCCACAGCGTGGCT 5820
Db 5761 TCTCTTCAACATATGGGGGGTGGTGGCTGGCCAGCTCGCCGCCCCCGGTGGCGCTA 5820
QY 5821 CAGCTTTCGTGGCGCGGCATCGCCGAGCGGCTGTGGCAGCATAGGCCCTTGGGAAGG 5880
Db 5821 CCGCCCTTCTGGCGCTGGCTTAGCTTGGCGCGCCATCGCGAGCTTGGACTGGGAAGG 5880
QY 5881 TCGTGTGACATCTTGGCGGCTATGGCGGAGGTAGCGCGGCACTCGTGGGCTTTA 5940
Db 5881 TCTGTGTGACATCTTGGCGGCTATGGCGGAGGTAGCGCGGCTTGTAGCATTC 5940
QY 5941 AGGTATGAGCGCGGAGGTGCCCTCAACGAGGAGCTGTCAACTTACTCCCTGCCATCC 6000
Db 5941 AGATCATGAGCGGTGAGTCCCTCAACGAGGAGCTGTCAACTTACTCCCTGCCATCC 6000
QY 6001 TCTCTCTGGTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6060
Db 6001 TCTCGCTTGGAGCCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6060
QY 6061 GCGCGGAGAGGAGCTGTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6120
Db 6061 GCGCGGAGAGGAGCTGTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6120
QY 6121 ACCAGCTTCCCTAGGCACTATGTGCTGAGAGGAGCTGACAGCACTGCTCAGT 6180
Db 6121 ACCATGTTTCCCTAGGCACTATGTGCTGAGAGGAGCTGACAGCACTGCTCAGT 6180
QY 6181 TCTCTCTAGCTTACCATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 6240
Db 6181 TACTCAGAGCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 6240
QY 6241 ACTGCTTACGCCATCTCCGCTCGTGGTGAAGGATGTTGGGATGGATGACAGG 6300
Db 6241 AGTGTACCACTCACTCCGCTCGTGGTGAAGGATGTTGGGATGGATGACAGG 6300
QY 6301 TGTGACTGACTTCAAGACCTGGCTCCAGTCCAACTCTGCGCGGTTTACCGGAGTCC 6360
Db 6301 TGTGAGCGACTTAAAGCTTGGCTGAAGCCAGCTCATGCCAACAATGCTGGGATTC 6360
QY 6361 CTTTCTCTATGCAACGCGGATCAAGGAGTCTGGCGGGGAGCGGATCATGCAAA 6420
Db 6361 CTTTGTGTCTGCGAGCGGCTATAGGGGGTCTGGCGAGGAGACGCAATTATGACA 6420
QY 6421 CCACCTGCCATGCGGAGCAGATGCGCGGACATGTCAAAACCGTTCCATGAGGATCG 6480
Db 6421 CTCGTGCACTGTGGAGCTGAGATCACTGGACATGTCAAAACCGGATGAGGATCG 6480
QY 6481 TAGGCCCTAGAACCTGAGCAACAGTGGCGAGGAGCTTCCCATCAACGATACACA 6540
Db 6481 TCGGTCTTAGGACCTGCGAGGAACATGTGGAGTGGAGCTTCCCATTAAGCCCTACACA 6540
QY 6541 CGGAGCTTGCACACCTCCCGCGGCCAATATCCAGGCGCTATGGCGGTTGGCTG 6600
Db 6541 CGGCGCTCTGACTCTCCCTCTCTGCGCGCACTATAAGTGTGGGCTGTGGAGGTTCTG 6600
QY 6601 CTGAGGAGTACTGTGAGGTTACGCTGTGGGGATTTCCACTACTGTGAGCGGATGACA 6660
Db 6601 CAGAGGAATATGAGGAGATAGCGGGTGGGGACTTCCACTACTGTGAGGATGACTA 6660
QY 6661 CTGACAAGTAAAGTCCCATGCCAGGTTCCGGCCCCCGAATTTCTTACGGAGGTGGATG 6720
Db 6661 CTGACAACTTAAATGCGCGTGCAGATCCCATCCCGCAATTTTCAAGAAATGGAGG 6720
QY 6721 GAGTGGTGGACAGGTACGCTCCGGGCTGCAAACTCTTCTTACGGGAGGAGCTACGT 6780
Db 6721 GGTGCGCTCATATAGGTTTGGCGCCCTTGCAGCCCTTGTGCGGGAGGAGGTATCAT 6780
QY 6781 TCCAGTCCGGCTCAACCACTACTGTGCTGGGTGCGAGCTCCCATGCGAGCCGACCGG 6840

Db 6781 TCAGAGTAGGACTCCACAGTACCCGGTGGGTGCCAATTACTTTCGAGCGCGAACC 6840
QY 6841 AGGTAAAGTGTACTTTCATCTCCATGCTCAGGATCCCTCCCATATACAGAGAGAGCGCTA 6900
Db 6841 AGTAAAGCGGTGTGAGCTCCATGCTCTCACTGATCCCTCCCATATATAACAGAGAGGCGCG 6900
QY 6901 AGCGTAGGCTGGCTAGAGGCTCTCCCTCTCTTACGAGCTCATCAGTACCGCAGTGTG 6960
Db 6901 GGAGAGGTTGGGAGAGGCTACCCCTCTTCTATGGCCAGCTCTTCGCGCAGCAGCTGT 6960
QY 6961 CTGCGCTCTTCTTGAAGGCGATGCACTAOCACCATGACTCCCGGAGCAGCTCA 7020
Db 6961 CCGTCTCATCTCTCAAGGCAACTTGCAACGCGCAACTGACTCTCCCTGACGCGAGCTCA 7020
QY 7021 TCGAGGCCAACTCTGTGGCGGAGGAGTGGGGGAGGAGTGGGGGAGGAGTGGGGGAGTGG 7080
Db 7021 TAGAGGCTAACTCTCTGTGGAGGAGGAGTGGGGGAGGAGTGGGGGAGGAGTGGGGGAG 7080
QY 7081 AGAATAAGGTAGTATTTCTGGACTCTTTTCCGAACGCTTCAACGCGAGGGGATGAGAGG 7140
Db 7081 AGAACAAAGTGTGATTTCTGGACTCTTTCGATCCGCTTGTGGCAGAGAGGATGAGCGGG 7140
QY 7141 AGATATCCGTGCGCGGAGATCTCGGAAATCCAGAAATTCAGAAATTCCTTACGCTTGGCCCA 7200
Db 7141 AGGTCTCGGTACCCCGCAGAAATTTCTGGGAAAGTCTGGAGATTCGCGGCGCTGCGCG 7200
QY 7201 TATGGGACGCGCGGACTTACATCTCCACTGTAGAGTCTTGGAGGAGCAGCGACTACG 7260
Db 7201 TTTGGGCGCGCGGACTTACAAACCCCGCTAGTAGAGAGCTGGAAGAGCTGACTACG 7260
QY 7261 TCCCTCTCGGTGTACAGGATGCCCATTTGCCACTTACCAAGCTTCCCAATACCACTC 7320
Db 7261 AACCACTGTGTTCATGGCTGCGCGCTACCACTTCCAGCTTCCCTCTCTGCTCGC 7320
QY 7321 CACGAGAAAGAGAGCGTGTCTGAGAGAGTCCAAATGCTCTTCTGCTTGGCGGAGC 7380
Db 7321 CTCGAAAAAGAGTACGCTGCTTCCCGAATCAACCTTACTTGTCTTGGCGGAGC 7380
QY 7381 TCGGCACATTAAGACCTTGGTGTCTCGGATCGCTGCGCGTGTAGAGGAGCAGCGGACCG 7440
Db 7381 TTGCCACAAAAGTTTGGCAGCTCTCAACTTCGGGCAATTAACGGGCGCAATATGACAA 7440
QY 7441 CCCTTCTGACTCGCTCCAGACGAGCTGCAAAAGGATCCGAGTGTGAGTCTGACTCTCT 7500
Db 7441 CATCTCTGAGCCCGCTTCTGCTGCTGCGCTGCGCGCTCCGAGCTTGGCTTATCTT 7500
QY 7501 CCATGCCCCCTTGAAGGGAGCGGGGAGCCCGGATCTCAGGACGCGGTCTTGGTCTA 7560
Db 7501 CCATGCCCCCTTGAAGGGAGCGGGGAGCTCGGGATTCGCGGATTCAGCGAGGCTCATGCT 7560
QY 7561 CCGT --- GAGTACGAGGCTAGTACGAGTGTGCTGCTCAATGTCTTATACGTGA 7617
Db 7561 CCGTACGAGTGGGGCGCAGCAGAGATGTGCTGCTGCTCAATGTCTTATACGTGA 7620
QY 7618 CAGCGCCCTGATCACCCCTATGCGCTCGGAGGAAAGTAAAGTTCGCGCATCAACCGTGA 7677
Db 7621 CAGCGCACTCTGCTCAACCCGCTGCGTGGGAGAAACAAACATGCCCATCAACGCACTGA 7680
QY 7678 GCAACTCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7737
Db 7681 GCAACTCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7740
QY 7738 TCCGCGCAAGAGGCTTACGAGTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7797
Db 7741 AAAGGAGAGAGAGTCACTTTCAGAGATGTGAGTGTGAGAGAGGATTCAGAGAGG 7800
QY 7798 TACTCAAGAGAGTGAAGCGGAGCGCTCCAGTAAAGGCTTAAGCTTCTATCTATAGG 7857
Db 7801 TGCCTCAAGGAGTCAAGAGCGCGCTCAAAAGTGAAGGCTAACTTCTATCCGTAGAGG 7860
QY 7858 AGGCTTCAAGTGAAGCGCGCTCACTTGGCGCAATCCAAATTTGGCTATGCGGCAAGG 7917

Db 7861 AAGCTTGACGCTGACGCGCCCGCCATTCAGCCAAATCCAAAGTTTGGCTATGGGCGCAAAAG 7920
Qy 7918 ACCTGCGGAACCTTATCAGCAGGCGCGTTAAACACATCCCGCTCCGCTGTGGAGACTTGC 7977
Db 7921 ACCTGCGTTGCCATGCGCAGAAAGCGGTAGCCACATCACTCCGCTGTGGAAAGACTTTC 7980
Qy 7978 TGAAGACACTGAAACACCAATTTGACACACCATCATGGAACAAAGAGAGTTTCTGCG 8037
Db 7981 TGAAGACACTGTAACACCAATTAAGACACTATCATCATGGAACAAAGAGAGTTTCTGCG 8040
Qy 8038 TCCAAACAGAGAGAGGAGCGCGCAAGCAGCTCGCTTATTCGATTTCCAGACCTCGGAG 8097
Db 8041 TCCAGCTGAGAGAGGAGGCTGTAAGCAGCTCGTATCATCGTGTTCCTCCGAGCTGGCG 8100
Qy 8098 TCCGCTGATGCGAGAGATGGCCCTTTAGACGTGCTTCCACACCTTCCCTCAGGCGGTGA 8157
Db 8101 TCGCGCTGTCGAGAGAGATGGCCCTGTACGAGCTGTAGCAAACTCCCTCCGCGCTGA 8160
Qy 8158 TGGGCTCCTCATACGATTTCAATCTCCCGCAGAGCGGGTGCAGTTCCTGGTGAATA 8217
Db 8161 TGGGAAGCTTCTACCGAATCCAAATCTACAGAGAGCGGGTGAATTCCTGCTCAAG 8220
Qy 8218 CTTGGAATCAAGAAATGCCCTATGGCTTCTCATATGACACCCCGCTGTTTGTCTCAA 8277
Db 8221 CTTGGAAGTCCAAAGAGACCCGATGGGTTCCCGTATGATACCCGCTGTTTGTCTCAA 8280
Qy 8278 CGTCACTGAGAGTGAATTCGTTGTGAGAGTCAATTTACCAATTTGTGACTTGGCC 8337
Db 8281 CAGTCACTGAGAGCGACATCCGTTACGAGAGAGCGAATTTACCAATTTGTGACTTGGCC 8340
Qy 8338 CCGAGGCGACAGAGCGCCATAAGGTCCTCAACAGAGCGGCTTTACATCGGGGTCCTGTA 8397
Db 8341 CCCAAGCCGCTGGCCATCAAGTCCCTCAGTGAAGGCTTTATGTTGGGGGCCCTCTTAA 8400
Qy 8398 CTAACCTAAAGGCGAGAACTCGGCTTATCGGCTGCGCGCAAGTGCCTGCTGACGA 8457
Db 8401 CCAATTAAGGGGGAAGAACTCGGCTATCGAGAGTCCGCGGAGCGGCTACTGACAA 8460
Qy 8458 CTAGTCGCGTAAATACCCCTCAATGTTTCTGAAGCCCACTGAGCCCTCGAGCTGCA 8517
Db 8461 CTAGTCGCGTAAATACCCCTCAATGTTTCTGAAGCCCACTGAGCCCTCGAGCTGCA 8520
Qy 8518 AGCTCAGGACTGACGATGCTGCTGAACGAGAGACGCTTCTGCTTATCTGTGAAGCG 8577
Db 8521 GACTCCAGGACTGACCATGCTGCTGTTGAGGAGCACTTATGCTGTAAGAGT 8580
Qy 8578 CCGGAACCCAGAGAGTGGCGGCTTACGAGCCCTTACGAGGCTATGATAGGTATT 8637
Db 8581 CCGGGGTCAGGAGAGCGCGGAGCTGAGAGCCCTTACGAGGCTATGACGAGTACT 8640
Qy 8638 CCGCCCCCGCGGGATCCGCCCAACAGAGATACGCTGAGCTGATGATGATGATGAT 8697
Db 8641 CCGCCCCCGCGGGACCCCGCACCAACAGATACGCTTGGAGCTTATGATGATGAT 8700
Qy 8698 CTTCAATGTTGCTGCTGCGGACGATGCTGCTGCAAGAGGCTATACCTACCTACCTG 8757
Db 8701 CTTCAATGTTGCTGCTGCGGACGCTGCTGCAAGAGGCTATACCTACCTACCTG 8760
Qy 8758 ACCCAACACCCCTTGTGACGCGGCTGCTGCAAGAGGCTATACCTACCTACCTG 8817
Db 8761 ACCCAACACCCCTTGTGACGCGGCTGCTGCAAGAGGCTATACCTACCTACCTG 8820
Qy 8818 CTTGCTAGGCAATATCATGATGCTGCTGCAAGAGGCTATACCTACCTACCTG 8877
Db 8821 CTTGCTAGGCAATATCATGATGCTGCTGCAAGAGGCTATACCTACCTACCTG 8880
Qy 8878 CTTGCTAGGCAATATCATGATGCTGCTGCAAGAGGCTATACCTACCTACCTG 8937
Db 8881 CTTGCTAGGCAATATCATGATGCTGCTGCAAGAGGCTATACCTACCTACCTG 8940
Qy 8938 TCTAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8997
Db 8941 TCTAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9000

Qy 8998 ATGCTCTTAGCGCATTTTACCTCCAGTTACTCTCCAGTGAGATCATATGGGTGGCTT 9057
Db 9001 ATGCTCTTAGCGCATTTTACCTCCAGTTACTCTCCAGTGAGATCATATGGGTGGCTT 9060
Qy 9058 CATGCTCAGGAAACTTGGGCTTACCCCTTTCGCAACCTTGGAGACATGGGCGCAGAAATG 9117
Db 9061 CATGCTCAGGAAACTTGGGCTTTCGCAACCTTTCGCAACCTTGGAGACATGGGCGCAGAAATG 9120
Qy 9118 TCGGCGCTAAGCTACTGTTCCAGGGGGGAGGCGCCACCTTGTGGGAGATACCTCTTTA 9177
Db 9121 TCGGCGCTAAGCTACTGTTCCAGGGGGGAGGCGCCACCTTGTGGGAGATACCTCTTTA 9180
Qy 9178 ACTGCGCAGTAAAGCAACGCTTAACTCACTCCAAATCCGCGCGCTCCAGCTGGACT 9237
Db 9181 ACTGCGCAGTAAAGCAACGCTTAACTCACTCCAAATCCGCGCGCTCCAGCTGGACT 9240
Qy 9238 TGTCTGGCTGGTTCGCTGCTGCTTACAGCGGGGAGACATATATCACAGCTGCTCTG 9297
Db 9241 TGTCTGGCTGGTTCGCTGCTTACAGCGGGGAGACATATATCACAGCTGCTCTG 9300
Qy 9298 CCGGACCCCGCTGGTTCGCTGCTTACAGCTCCAAATCCGCGCGCTCCAGCTGGACT 9357
Db 9301 CCGGACCCCGCTGGTTCGCTGCTTACAGCTCCAAATCCGCGCGCTCCAGCTGGACT 9360
Qy 9358 TGTCTGGCTGGTTCGCTGCTTACAGCTCCAAATCCGCGCGCTCCAGCTGGACT 9411
Db 9361 TGTCTGGCTGGTTCGCTGCTTACAGCTCCAAATCCGCGCGCTCCAGCTGGACT 9414

RESULT 15

US-09-995-937-19
; Sequence 19, Application US/0995937
; Publication No. US20030028010A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995,937
; FILING DATE: 28-Nov-95 US20030028010A1-2001
; CLASSIFICATION: <unknown>
; APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 19;
US-09-995-937-19

Query Match 63.9%; Score 6135.6; DB 10; Length 9416;
Best Local Similarity 78.3%; Pred. No. 0;
Matches 7372; Conservative 0; Mismatches 2039; Indels 3; Gaps 1;

QY	1	GCAGCCCTCTGATGGGGGACACTCCACATCAATCACTCCCTCTGAGGAACACTG	60	Db	961	GCCTTAATTCGAGTATTGTGTACGAGGGCCGATGCCATCTCTGACACTCCGGGGGTG	1020
Db	1	GCAGCCCTCTGATGGGGGACACTCCACATCAATCACTCCCTCTGAGGAACACTG	60	QY	1021	TGCTCTGTGTTCAGGAGGTAAACAGCTCCCTGTGTGGTAGCGCTCACTCCACGCTCG	1080
QY	61	TCTTCAGGCAAGAGCTGTAGCATGGCTTGTAGTATGATGTCTGTCAGCTCCAGGAC	120	Db	1021	TCCCTTGGCTTCGAGGAGGTAAACGCTTCGAGGTGTGGTGGCGGTGACCCCAACGCTG	1080
Db	61	TCTTCAGGCAAGAGCTGTAGCATGGCTTGTAGTATGATGTCTGTCAGCTCCAGGAC	120	QY	1081	CGGCCAGGAATGCCAGGTCCCACTACACATACAGACGCCACGTGCTGCTGCTG	1140
QY	121	CCCCCTCCCGGAGAGCCATAGTGTCTCGGAAACGGTGTAGTACACCGAATGCCAG	180	Db	1081	CMACAGGACGGCAAACTCCCAACACGACGCTTCGACGTATATCGATCTGCTGTG	1140
Db	121	CCCCCTCCCGGAGAGCCATAGTGTCTCGGAAACGGTGTAGTACACCGAATGCCAG	180	QY	1141	GGACGCTCTTCTCTCTGCTCGGTATGTACGTGGGGATCTCTGCGGATCTATTTTCTCTG	1200
QY	181	GACGACGGGTCTTTCTTTGGATCAACCCCTCAATGCCCTGGAGATTGGGCGTGCCTCC	240	Db	1141	GGAGCGCACCTCTGCTCAGCCCTCTAGTGGGAGACCTGTGCGGTCTGTTTTCTTG	1200
Db	181	GACGACGGGTCTTTCTTTGGATCAACCCCTCAATGCCCTGGAGATTGGGCGTGCCTCC	240	QY	1201	TCCTCCAGCTGTTCACCTTCTGCTCCGCGGCAATGAGACAGTGTGCAAGTCTCAACTGT	1260
QY	241	GGAGACTCTAGCCAGTAGTGTGGTTCGGAAGAGCTTGTGTGCTGCTGCTGATAGG	300	Db	1201	TTGGTCAACTGTCTTCTCTCCAGCGGCACTGGACGACGCAAGCTGCAATGT	1260
Db	241	GAAGACTCTAGCCAGTAGTGTGGTTCGGAAGAGCTTGTGTGCTGCTGCTGATAGG	300	QY	1261	CAATCTATCCCGGCCATGTATCAGGTCAACGCTCCGATCCCAAGCTGCTGTGACA	1320
QY	301	GTGCTTGCAGTCCCGGAGGTCTCGTAGACCGTGCACCATGACGACGATCTTAAC	360	Db	1261	CTATCTATCCCGGCCATGTATCAGGTCAACGCTCCGATGGGATATGATGATGAACTGT	1320
Db	301	GTGCTTGCAGTCCCGGAGGTCTCGTAGACCGTGCACCATGACGACGATCTTAAC	360	QY	1321	CACCTAACACAGCCCTAGTGTGTGCTGCTCCGATCCCAAGCTGCTGTGACA	1380
QY	361	CTCAAGAAAAACCAACCTGTACCAACACCGCCGCCACAGGACGTCAGTTCCTCGGGG	420	Db	1321	CCCTAGGACAGCTGTGTGTGTAGTCTCAGTCTCGGTCCGATCCCAAGCTGCTGTGACA	1380
Db	361	CTCAAGAAAAACCAACCTGTACCAACACCGCCGCCACAGGACGTCAGTTCCTCGGGG	420	QY	1381	TGGTGGCGGGGCGCACCTGGGGAGTCTGGCGGGCTTGCCTACTATTCATGTGTAGGA	1440
QY	421	GTGCTCAGATCGTGTGTGAGTTTACTGTTGCGGCGCAGGGGCCCAAGTGTGGTGTCC	480	Db	1381	TGATCGCTGCTCACTGGGAGTCTGGCGGGCATAGCGTATTTCTCCATGTGTGGGA	1440
Db	421	GTGCTCAGATCGTGTGTGAGTTTACTGTTGCGGCGCAGGGGCCCAAGTGTGGTGTCC	480	QY	1441	ACTGSGCTAAGTCTGTGATGTGGCTACTCTTTGCGGGCTTGAACGGGAGACCCACA	1500
QY	481	GCGGACTAGGAAGGTTCCGAGCGTGCACACCTGTCGMAAGGCGACACCTATCCCAA	540	Db	1441	ACTGSGCGAGGTCTGTGTGTGTGCTGCTATTTGCGGGCTGACGCGGAAACCCAG	1500
Db	481	GCGGACTAGGAAGGTTCCGAGCGTGCACACCTGTCGMAAGGCGACACCTATCCCAA	540	QY	1501	CGACGGGAGGTGGCGGCCACACACCTCCGGGTTCACGCTCTTCTCATCTGCGG	1560
QY	541	AGGCTGCGCGACCCGAGGCGAGGCTGGCTCAGCCCGGTACCTTGGCCCTCTATG	600	Db	1501	TCACCGGGGAGTTCGCGGCCACACACCGCTGGGCTTGTGGTCTCTCTTACACAGCG	1560
Db	541	AGGCTGCGCGACCCGAGGCGAGGCTGGCTCAGCCCGGTACCTTGGCCCTCTATG	600	QY	1561	CGTCTCAGAAATCAGCTTGTGAATAACACAGGAGTGGCACATCAACAGGACTGCC	1620
QY	601	GCAATGAGGCTGGGTTGGGAGGAGTCTCTGTCAACCGCGCTCCCGCTAGTT	660	Db	1561	CCAAGCAGAAATCACTCACTGATCAACACAGGAGTGGCACATCAACAGGACTGCC	1620
Db	601	GCAATGAGGCTGGGTTGGGAGGAGTCTCTGTCAACCGCGCTCCCGCTAGTT	660	QY	1621	TAAATGTCAATGACTCCCTCCAACTGCTGCTTCTTTGCGGCTTGTTCACGACAGT	1680
QY	661	GGGCCCCACGACCCCGCGGTAGTTCGCTAACTTGGGTAAAGTCACTCGATACCTTA	720	Db	1621	TCACTGTCAAGTAGCTTACCCCGCTGGTAGAGGGCTCTTCTATGCCCAAAAT	1680
Db	661	GGGCCCCACGACCCCGCGGTAGTTCGCTAACTTGGGTAAAGTCACTCGATACCTTA	720	QY	1681	TCAACTCTCAGGCTGTCTGAGAGGTTGGCAGCTGCCGAGCTTACCGATTTTGCC	1740
QY	721	CATGGGCTTCCGCTATGAGGTACATTCGCTCGCGCGCCCTTAGGGGCG	780	Db	1681	AGGGTGGGGTCCCATCATGCTTATGCCAAGGAGCGGCTTGAACAGCGCCCTACTGT	1800
Db	721	CATGGGCTTCCGCTATGAGGTACATTCGCTCGCGCGCCCTTAGGGGCG	780	QY	1741	GGCATTAGCGCTCGACCGTGTGTGTCTACCGCTGCGAGTTCAGAGGCTTATGCT	1860
QY	781	CTGCCAGGCTTGGACACAGGTTCGGGTCTGGAGACCGGCTGAATGAAACAG	840	Db	1741	GGCATTAGCGCTCGACCGTGTGTGTCTACCGCTGCGAGTTCAGAGGCTTATGCT	1860
Db	781	CTGCCAGGCTTGGACACAGGTTCGGGTCTGGAGACCGGCTGAATGAAACAG	840	QY	1801	GGCATTAGCGCTCGACCGTGTGTGTCTACCGCTGCGAGTTCAGAGGCTTATGCT	1860
QY	841	GGAACTTCCCGGTGCTCTTCTATCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTG	900	Db	1801	GGCATTAGCGCTCGACCGTGTGTGTCTACCGCTGCGAGTTCAGAGGCTTATGCT	1860
Db	841	GGAACTTCCCGGTGCTCTTCTATCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTG	900	QY	1861	ATGTTTCAACCCCAAGCCCTTGTGTGGGGAACACCGATGCTTCCGCTTCCCTACGT	1920
QY	901	TCCAGCTTCCGCTATGAGGTGCGCAACGTCGCGGATATACCATGTCAAGAACGACT	960	Db	1861	ATGTTTCAACCCCAAGCCCTTGTGTGGGGAACACCGATGCTTCCGCTTCCCTACGT	1920
Db	901	TCCAGCTTCCGCTATGAGGTGCGCAACGTCGCGGATATACCATGTCAAGAACGACT	960	QY	1921	ATAGTGGGGGAGATGAGACAGAGTGTCTCTCTCAACACCGGCTCCGCAACAG	1980
QY	961	GCTCAACTCAGCAATGTGATAGGAGCGGATGATGATGATGATGATGATGATGATG	1020	Db	1921	ATAGTGGGGGAGATGAGACAGAGTGTCTCTCTCAACACCGGCTCCGCAACAG	1980
Db	961	GCTCAACTCAGCAATGTGATAGGAGCGGATGATGATGATGATGATGATGATGATG	1020	QY	1981	CAACTGTGCTCGGTGTACATGGATGATGATGATGATGATGATGATGATGATGATG	2040
QY	1021	GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1080	Db	1981	CAACTGTGCTCGGTGTACATGGATGATGATGATGATGATGATGATGATGATGATG	2040
Db	1021	GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1080	QY	2041	CCCGTGTAAATCGGGGGGTGCTTAAACCGACCTTGTATCTGCCCAACGACTGCTTCC	2100
QY	1081	GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1140	Db	2041	CCCGTGTGTATCGGGGGGTGCGCAACACCTTGTCTGCTGCTGCTGCTGCTGCTTCC	2100
Db	1081	GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1140				

2101 GGAAGCACCCCGAGGCTACTTACACAAATGTGGCTCGGGCCCTGGTTGACACCTTAGGT 2160
2101 GCAAAATCGGAAGCACATACTCTCGGTGGGCTCCGGTCCCTGGATTACACCCAGGT 2160
2161 GCCTAGTAGACTACCCATACAGGCTTTGGCACTACCCCTGCATCTCAATTTTCCATCT 2220
2161 GCATGGTGAATAACCCGTATAGGCTTTGGCACTATCTTTGTACTATCAATTTACACCATAT 2220
2221 TTAAGGTTAGGATGTATGTGGGGGGGTGAGACACAGGCTCAATGCCGATGCAATTTGA 2280
2221 TCAAGGTACAGATGTACGTGGAGGGTCCAGACAGGCTGGAACGGCTGGAATCTGGA 2280
2281 CTCGAGAGAGCCTGTAACTTTGGAGACAGGATAGGTAGAACTCAGCCCGCTCTGC 2340
2281 CGCGGGCGAAGCCTGTGTATCTGGAAGACAGGACAGGTCGAGCTCAGCCCATCTCTCC 2340
2341 TGTCTAAACAGAGTGCAGATACCTCCCTGTCTTCTACACCTTACCGGCTTTATCCA 2400
2341 TGTCCACACACAGTGGCAGGTCCTTCCGTGTTCTTTTCAAGCCCTGCCAGCCTTTGCCA 2400
2401 CTGGTTTGTATCCA TCTCCATCAGAACTCGTGACGTGCAATACCTGTACGGTGTAGGCT 2460
2401 CCGGCTCATCCACTCCACAGAACTTGTGACGTGCAGTACTTGTACGGGTGGGT 2460
2461 CAGCGTTTGTCTCTTGTGCAATCAAAATGGAGTACATCTGTGTGTTTCTCTCTCTG 2520
2461 CAAGCATCGCTCTGGGGCCATTAAGTGGAGTACCTGCTCTCTCTCTCTCTCTCTG 2520
2521 CAGACCGCGGTGTGCTGCTGTGTGATGATGCTGTGATAGCCAGCCAGCTGAGCCG 2580
2521 CAGACCGCGGTGTGCTGCTGTGTGATGATGCTGTGATAGCCAGCCAGCTGAGCCG 2580
2581 CTTTGAAGAACTGGTAATCTCAATGACAGATCTCTGCGGAGCGCATGATATCTCT 2640
2581 CTTTGAAGAACTGGTAATCTCAATGACAGATCTCTGCGGAGCGCATGATATCTCT 2640
2641 CTTTCTTGTGTCTCTCTGCGGCTCTGTGATCATTTAAGGCAAGCTGCTCTCTGCGGCG 2700
2641 CTTTCTCTGTGTCTCTCTGCGGCTCTGTGATCATTTAAGGCAAGCTGCTCTCTGCGGCG 2700
2701 CGTATGCTTTTATGGCGTATGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
2701 TCTACGCTTCTACGGGATGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
2761 CTTTACGCTTGTGACCGGAGATGTGTGATCTGCGGGGTGCGGTCTCTGTAGTCTGTG 2820
2761 CATACGCACTGGACACGGAGTGGCGCGTGTGTGTGGCGGTGTGTCTTGTCTGTGAA 2820
2821 TATCTTGTACCTTGTACCACTACTACAAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
2821 TGGCGTGACTCTGTACCACTATACAGCGCTATATCAGCTGTGTGTGTGTGTGTGTGT 2880
2881 AATACTTTTATACAGAGCGGAGGCGCATGCAAGTGTGGTGTGGTGTGGTGTGGTGTGG 2940
2881 AGTATTTCTGACAGAGTAGAGGCGCACTGCAAGTGTGGTGTGGTGTGGTGTGGTGTGG 2940
2941 GGGAGGCGGCGATGCGATFCT 3000
2941 GGGAGGCGGCGATGCGGCT 3000
3001 ACATCACCAAACT 3060
3001 ACATCACCAAACT 3060
3061 CGAGATGCGGTACT 3120
3061 TTAAGTCCCTTACT 3120
3121 AAGTCCGCGGGGTCTATTTGTCAAATGGTCTTCAAGAGCTGGCGGCTGTGACAGGTA 3180
3121 AGATAGCGGGAGTCTATTTACGTGCAATGGCCATCATCAAGTTGGGGGCGCTTACTGGCA 3180

3181 CCTAGTTTATTAACCATCTTACCCACTGCGGCACTGGGCGCACCGGGCCCTACGAGAC 3240
3181 CCTATGTGTATTAACCATCTTACCCACTGCGGCACTGGGCGCACCGGGCCCTACGAGATC 3240
3241 TTGCGGTGGCGGTAGAGCCCGTCTCTCTCTCGCCATCGAGACCAAGGTCTATCACTCTGG 3300
3241 TGGCGGTGCTGTGGAACCAAGTCTCTCTCTCCGAATGAGACCAAGGTCTATCACTCTGG 3300
3301 GAGCAGACACCGCTCGGTGTGGGACATCACTCTGGGTCTACCGTCTCTCGCCCGCAAGG 3360
3301 GGGCAGATACCCCGCGTGGGTGACATCATCAACGGGTCTGCGCGTCTCTGCGCGTAGG 3360
3361 GGAAGGAGATATTTTGGGACCGGCTGATGTCTCGAAGGCAAGGGTGGGACTCTCTTG 3420
3361 GCCAGGAGATCTGCTTGGACCGCGGAAATGTCTCTCAAGGGTGGAGGTGTCTGG 3420
3421 GCGCCATCAGGCTACTCTCCAAACGCGGGGTACTTGGTGTGATCATCTACATGCC 3480
3421 GCGCCATCAGGCTACTCTCCAAACGCGGAAATGTCTCTCAAGGGTGGAGGTGTCTGG 3480
3481 TCACAGCGCGGACAAAGAACAGGTCGAAAGGGAGGTTCAAGTGGTCTTACCCCAACAC 3540
3481 TCACTGGCGGACAAACAAAGTGGAGGTGAGGTCCAGATCTGTCTCACTGTCTACCT 3540
3541 AATCTTCTCTGGCACTCTCATCAACGCGGTGTCTGACTGTCTTACCATGCGCGTGGCT 3600
3541 AAACCTTCTGCGCAACGTGATCAATGGGTATGCTGACTGTCTACCAAGGGCCGAA 3600
3601 CGAGACCTAGCCTGCTCAAAAGGTCAATCAACCAAAATGTACCAATGTAGACCTGG 3660
3601 CGAGACCTAGCCTGCTCAAAAGGTCAATCAACCAAAATGTACCAATGTAGACCTGG 3660
3661 ACCTGCTGCGGTGGCAGCGCGCCCGGGCGGCTCCATGACACCACTGACGCTGTGCA 3720
3661 ACCTTGTGGGTGGCGGCTCTCTCAAGTTCCTGCTCAATGACACCTGTGACCTGCGGCT 3720
3721 GCTCGACCTTTACTTGGTCAAGAGATGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 3780
3721 CTTGCGACCTTTACTCTGTTACGAGGACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3780
3781 ACAGCAGGGAGTCTACT 3840
3781 ATAGCAGGGTAGCTGCT 3840
3841 GTCAATCT 3900
3841 GTCTGCTGTCT 3900
3901 GGGGGTGGCGAAGCGGTGCACTTCTATACCGTGTGATCTCTATGAACTACCACTGCGGT 3960
3901 GTGAGTGAACCAAGCGGTGGACTTTATCTCTGTGGAGAACCTAGAGACAACTATGAT 3960
3961 CTCTGCTTCTACAGACACTCAACCCCGCGCTGTACCGCAGACATTCACAGTGGCAC 4020
3961 CCGCGGTCTTACCGACACTCTCTCTCTCAACAGAGTGGCCAGAGCTTCCAGTGGCCCC 4020
4021 ATCTGCACT 4080
4021 ACCTGCACT 4080
4081 AAGGTTACAGGTCT 4140
4081 AGGCTCAAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4140
4141 ATATGTCACAGGACAGGTTATCGACCTTACATCAGAACTGGGGTGAAGACATTTACCA 4200
4141 ACATGTCACAGGACAGGTTATCGACCTTACATCAGAACTGGGGTGAAGACATTTACCA 4200
4201 CCGCGGTCT 4260
4201 CTGGCAGCCCATCACTGCT 4260
4261 GGGCGCTCTATGACATCATTAATATGTGATGATGATGATGATGATGATGATGATGATGAT 4320

Db 4261 GAGGGCTTATGACATATATTTTGTGACGAGTCCCACTCACGGATGCGACATCCATCT 4320
 Qy 4321 TGGGCATCGGCACAGTCTCGACCAAGCGGAGAGCGGTGGAGCGGCGTCTGTCGTGCTG 4380
 Db 4321 TGGGCATCGGCACAGTCTCGACCAAGCGGAGAGCGGTGGAGCGGCGTCTGTCGTGCTG 4380
 Qy 4381 CCACGGCTACACCTCCGGATCGGTACCGGTGCCACACCCCAATATCGAGGAATAGGCC 4440
 Db 4381 CCACGTCTACCCCTCCGGCTCCGGTCACTGTGTCCATCCCTAACATCGAGGAGTTGGTC 4440
 Qy 4441 TGTCCAAACAATGGAGAGATCCCTTCTATGGCAAGCCATCCCATTTAGGCCCATCAAGG 4500
 Db 4441 TGTCCAAACAAGAGATCCCTTCTATGGCAAGCCATCCCATTTAGGCCCATCAAGG 4500
 Qy 4501 GGGGAGGCACTCATTTTTGTGCGATTCGAGAGAAATGTGACGAGCTCCCGCAAGC 4560
 Db 4501 GGGGAAGACATCTCATCTTCTGCTCACTCAAGAAGAGTGCGAGCTCGCGCGAAGC 4560
 Qy 4561 TGACAGGCTCGGACTGAACCTGTAGCATATTAACGGGGCTTCATGTCCGTCATAC 4620
 Db 4561 TGCTCGCATTTGGGCATCAATGCGGTGGCTACTACCGGAGCTTGAAGTGTCTGATCC 4620
 Qy 4621 CCGCTATCGGAGACGTCTGTGCGTGGCAACAGAGCTCTAATGACGGGTTCACCGGC 4680
 Db 4621 CGACCAACGGCGATGTTGTGCTGTGTGCGACCGATGCTCTCATGACTGGCTTACCGGC 4680
 Qy 4681 ATTTGACTGATGATGACATGCAATACATGTTGTCAACCGAGCATGTGACCTTGG 4740
 Db 4681 ACTTCGACTCTGTGATAGATGCAACACGTTGTCTCACTCAGACATCGATTCAGCCTTG 4740
 Qy 4741 ATCCACACTTACCATTTAGAGACAGACCGTSCCCCAAGACCGGTGTCCGCTCGCAAC 4800
 Db 4741 ACCCTACCTTTACCATTTAGAGACACACGCTCCCAAGATGCTGTCTCCAGACTCAGC 4800
 Qy 4801 GCGAGGTAGAACTGGCAGGGTAGAGTGGCACTTACAGTGTGTGTAUCCAGAGAAC 4860
 Db 4801 GCGGGGAGGACTGCGAGGGGAGAGCGAGGATCTACAGATTTGTGGCACCGGGGAGC 4860
 Qy 4861 GCGCCTCGGGCAATGTTGGATTTCTGCTGTGAGTGTCTATGACGGGCTGTGCTT 4920
 Db 4861 GCGCCTCGGGCAATGTTGGATTTCTGCTGTGAGTGTCTATGACGGGCTGTGCTT 4920
 Qy 4921 GGTATGAGCTACGCGCTGAGACCTCGTGTAGTGTGGGCTTACCTAAATACACAG 4980
 Db 4921 GGTATGAGCTATGCGCGCGAGACTACAGTTAGGCTACAGGCTACATGACACCCCG 4980
 Qy 4981 GGTGCGGCTTSCCAGGACCAATCTGGAGTTCTGGGAGAGCGTCTTCAAGGCTCAACC 5040
 Db 4981 GGTGCGGCTTSCCAGGACCAATCTGGAGTTCTGGGAGAGCGTCTTCAAGGCTCAACC 5040
 Qy 5041 ACATAGATGCCACTTCTCTGCCAGACTAAACAGCAGAGACACTTTCCTTACCTGG 5100
 Db 5041 ATATAGATGCCACTTCTCTGCCAGACTAAACAGCAGAGACTTTCCTTACCTGG 5100
 Qy 5101 TGGCATATCAAGCTACAGTGTGGCGGACAGGCTCAAGCTCAACCTCCATCGTGGACCAA 5160
 Db 5101 TAGCGTACCAAGCACCGTGTGGCTTAGGGCTCAAGCCCTCCCCCATCTGTGGACCA 5160
 Qy 5161 TGTGGAAGTGTCTCATACGCTGAACCTTACTGACCGGGCAACACCCCTGCTGTATA 5220
 Db 5161 TGTGGAAGTGTCTCATACGCTGAACCTTACTGACCGGGCAACACCCCTGCTGTATA 5220
 Qy 5221 GGCTAGAGCGCTCAAATAGAGTCACTCTACACACCCCAATCAATTAATATCATG 5280
 Db 5221 GACTGGCGCTGTTCAGATGAGTCACTCTGACCGGCAACCAATATCATG 5280
 Qy 5281 CATGCATGCGGTGACCTGGAGTCTGCTACTAGCACTGGGTGCTGGTAGGCGGAGTCC 5340
 Db 5281 CATGCATGCGGTGACCTGGAGTCTGCTACTAGCACTGGGTGCTGGTAGGCGGAGTCC 5340
 Qy 5341 TTGCAGCTTTGGCGGATACTGCTGACGACAGGCACTGTGGTCACTTTGGGCGAGATCA 5400

Db 5341 TGCGTCTCGCGCGCTATTGCTGTCTCAACAGCTCGCTGGTCACTAGTGGGCGAGGATG 5400
 Qy 5401 TCTTGTCCGGAGAGCGAGCTGTGCTTCCGACAGGGAAGTCTCTACAGAGATTCGATG 5460
 Db 5401 TCTTGTCCGGAGAGCGGCAATTAATCTGACAGGAGGTTCTCTACAGAGATTCGATG 5460
 Qy 5461 AGATGGAAGAGTGTGCTCTCAACATTCCTTCTATCATCGAGAGGGAATGCACTGCGCCGAGC 5520
 Db 5461 AGATGGAAGAGTGTCTCTCAACATTCCTTCTATCATCGAGAGGGAATGATCTGCTGAGC 5520
 Qy 5521 AATTCAGAGAAAAGCGCTCGGTGTGTGCAAGGSCCAACAGAGCGAGGAGCTGCTG 5580
 Db 5521 AGTTCAGAGAAAAGCGCTCGGTCTCTGACAGCGGCTCCCGCATGACAGAGTTATCA 5580
 Qy 5581 CTCCGCTGTGTGAGTCCCAAGTGGCGAGCCCTTGTGAGACCTTCTGGGCGAAGCAATGTGGA 5640
 Db 5581 CCCCTGTGTCCAGACCACTGGGAGAAAATCTGAGGTTCTTCTGGGCGAAGCAATGTGGA 5640
 Qy 5641 ATTTTCATCAGCGGAATACAGTACCTAGAGCTTTATCACTCTCTCTGGAACCCCGGGA 5700
 Db 5641 ATTTTCATCAGTGGGATACATATTTGGCGGCTGTCAAAGCTGCTGTGTAAACCCGCA 5700
 Qy 5701 TAGCATATATGATGCAATTTACAGCTTTATCACTAGCCGCTCACACCCCTCACACCCAAACACCC 5760
 Db 5701 TTGCTTCATTTGATGGCTTTTACAGCTGCGGTCAACAGCCCACTAACCACTGCGCCAAACCC 5760
 Qy 5761 TCCTGTTTAAACATCTTGGGGGATGGGTGGTGTGCCCACTCGCTCTCCACAGCGTGGCT 5820
 Db 5761 TCCTTCTTCAACATTTGGGGGGTGGGTGGTGTGCCAGCTCGCCGCCCGGTGCGCTA 5820
 Qy 5821 CAGCTTTCTGTTGGCGCGCGCATCGCGAGCGGTGTGTGGCAGCATAGGCCCTTGGAGG 5880
 Db 5821 CCGCTTTGTTGGCGCTGGCTTAGCTGGCGCGCATCGGAGCGTTGGAATGCGGAGG 5880
 Qy 5881 TGCTCGTGGACATCTTGGCGGCTATGGGSCAGGGGTAGCCGCGCATCTGCTGGCTTTA 5940
 Db 5881 TCCTCGTGGACATCTTGGAGGTTATGGCGCGGCTGCGCGGAGCTCTTGTAGATTTCA 5940
 Qy 5941 AGCTCATCAGCGCGGAGTGCCTCCACCGAGGACCTGTCTCAATTAATCTCTGTCGATCC 6000
 Db 5941 AGATCATGAGCGGTGAGTGCCTCCACAGGAGACCTGTCTCAATCTGCTGCGCCATCC 6000
 Qy 6001 TCTCTCTGTGTGCTGCTGTGCGGGTGTGTGTGGCAGAGAACTGTGCTGCGGAGCTG 6060
 Db 6001 TCTGCTCTGAGCGCTTGTAGTGTGTGTGTGTGTGGCAGCAATTAATGCGCGGAGCTG 6060
 Qy 6061 GCGCGGAGAGGGGCTGTGAGTGGATGAACCGCTCATAGCTTGTGCTTCCGCGGCTA 6120
 Db 6061 GCGCGGAGAGGGGCTGTGAGTGGATGAACCGCTCATAGCTTGTGCTTCCGCGGCTA 6120
 Qy 6121 ACCAGCTCTCCCTACGCACTATGTGCTGAGAGCGAGCTGCAGACGTGTCACTCAGA 6180
 Db 6121 ACCATGTTTCCCCACGCACTAGCTGCGGAGAGGATGACAGCGCGCGCTCACTGCCA 6180
 Qy 6181 TCCTCTTAGCTTACCATCTCACTAACTGTGTAAGCGGCTCCACAGTGGATTAATGAGG 6240
 Db 6181 TACTCAGAGCGCTCACTGTAAACCGAGCTCTGAGCGCACTACATCACTGTAAGCTCGG 6240
 Qy 6241 ACTGCTTACGCACTGCTCCGCTCGTGTGTAAGGATGTTTGGATTTGGATATGCAAG 6300
 Db 6241 AGTGTACCACTCACTGCTCCGCTCTCTGCTAAGGGAATCTGGAGCTGATATCGAGG 6300
 Qy 6301 TCTTACTGACTCACTCAAGACCTGGCTCCAGTCCAAACTCTGCGCGGTTTACCGGAGTCC 6360
 Db 6301 TGCTGAGCGCATTTTGAACCTGGCTGAAGCAAGCTCATGCCCCAATCTGCTGGGATTC 6360
 Qy 6361 CTTTCTGTCTGTAAGCAAGCGGTTTCAAGGAGTGTGGCGGGGGAACGCACTCATGCAAA 6420
 Db 6361 CTTTGTGTCTGCGAGCGGGTATAGGGGGTCTGGCGAGAGACGCGCATTTATGCA 6420
 Qy 6421 CCACTGCCCCACGCGAGCACAGTCCGCGGATGTCAAAAACGTTTCCATGAGGATCG 6480
 Db 6421 CTCGCTGCCACTGTGAGCTGAGATCACTGGACATGTCAAAAACGCGAGATGAGGATCG 6480

QY 6481 TAGGGCTAGAACCTGAGCAACAGCTGGGACGGAACGTTCCCATCAACGCAATACACCA 6540
Db 6481 TCGTCTCTAGACCTGACGAAACATGTGAGTGGGACGTTCCCATTAACGCTACACCA 6540
QY 6541 CGGACCTTGCACACCTCCCGCGCCCAACTATTCCAGGGGCTATGGCGGGTGGCTG 6600
Db 6541 CGGACCTTGTACTCTCCCTTCTCGGCGCAACTATAAGTTCCGCGTGGAGGGTGTCTG 6600
QY 6601 CTGAGGAGTACGTGGAGGTACGGTGTGGGGATTTCCACTAGCTGACGGGCAATGACCA 6660
Db 6601 CAGAGGAATACGTGGAGATAAGGGGGTGGGGACTTCCACTAGTATCGGGTATGACTA 6660
QY 6661 CTGACAGTAAAGTGGCCATGCGAGTTCCGGCCCCCGGATTTCTTCAAGAGTGGATG 6720
Db 6661 CTGACATCTTAATGCGCGTGGCAGATCCCATCGCCGCAATTTTACAGAAATGGAGC 6720
QY 6721 GAGTGGGTTGCA CAGGTACGCTCCGGCGTCAAAACCTCTTCTACGGGAGGAGCTCAGT 6780
Db 6721 GGGTGGCCCTACATAGGTTTGGCGCCCTTGCAGGCCCTTCTCGGGAGGAGGTATCAT 6780
QY 6781 TCCAGGTCGGGCTCAACGATATCTTGGTGGGTGCGAGTCCCATGCGAGCCGAAACCGG 6840
Db 6781 TCAGAGTAGACTCCACAGTACCGGTGGGGTGGCAATTAACCTTGGAGCCCGAACCGG 6840
QY 6841 ACCTAACAGTGTCTTACTTCCATGCTCACCGATCCCTCCCATTTACAGCAGAGAGCGCTA 6900
Db 6841 ACCTAGCGGTGTGACGTCATGCTCTACTGATCCCTCCCATATAACAGCAGAGCGCGCG 6900
QY 6901 AGGTAGGCTGGCTAGAGGTCTCCCGCTCTTTAGCCAGCTCATCAGTAGCCAGTTGT 6960
Db 6901 GGAGAGGTTGGGAGAGGTCAACCCCTTCTATGGCCAGCTCTCGGCCAGCCAGCTGT 6960
QY 6961 CTCGCGCTCTTTGAGGGCAGCTGCACTACCAACGATCTCCCGGAGCGCTGACCTCA 7020
Db 6961 CCGCTCATCTCTCAGAGCAACTTGACCGCCCAACCATGACTCCCTCTGACGCGAGCTCA 7020
QY 7021 TCAGGCGCAACCTCTTGTGGCGGAGAGATGGCGGAACATCACTCTCGCTGGAGTCA 7080
Db 7021 TAGAGGCTAACCTCTGTGGAGCAGAGATGGCGGCAACATCA CAGGGTTGAGTCA 7080
QY 7081 AGAATAGGTAGTAAATTTGGACTCTTTGAAACCGCTTTCAGCGGAGGGGATGAGAGG 7140
Db 7081 AGAACAAAGTGGTGAATTTGGACTCTTCGATCCGCTTGTGGCAGAGGATGAGCGGG 7140
QY 7141 AGATATCCGTCGCGCGGAGATCTCGGAAATTCAGAGATTTCCCTCAGCGTTGCCA 7200
Db 7141 AGGTCTCCGTA CCGCGAGAAATTCGCGAAGTCTCGGAGATTCGCGCGGCGCTGCCG 7200
QY 7201 TATGGCACCGCGGACTCAATTCCTCCACTGTAGAGTCTCTGGAAGGCGGATGAGAGG 7260
Db 7201 TTTGGCGCGCGGACTACACCCCGCTAGTAGAGAGTGGAAAGGCTGACTAG 7260
QY 7261 TCCCTCCGTTGTA CAGGATGCCANTGCCACTACCAAGGCTCTCCAATACCACTTC 7320
Db 7261 AACCACTGTGTCCATGGCTGCCCTCAACCTCCACGGTCCCTCTCTGTGCTCCG 7320
QY 7321 CACGGAGAAAGGAGGACGGTTGCTGACAGAAATCCAAATGTTCTTCTGCTTGGCGAGC 7380
Db 7321 CTCGGAAAGGATCAGGTGTGCTTCAACGAATCAACCTACTGCTTGGCGGAGC 7380
QY 7381 TCGCCACTAAGACCTTCGGTAGCTCGGATCTCGGCGGTTGATAGCGGACGCGGACCG 7440
Db 7381 TTGCCACCAAAAGTTTGGCAGTCTCTCACTTCCGGCAATAGGGGCAATATAGCAA 7440
QY 7441 CCGTCTGACCTGGCTCCGAGCGGTGA CAAAGGATTCGACGTTGAGTGTACTCTCT 7500
Db 7441 CATCTCTGAGCGCGCCCTTCTGGGTGCCCCCGGACTCCGAGCTTGAGTCTTATCTT 7500
QY 7501 CCATGCCCCCTTGAAGGGAGCGGGGACCGGATCTCAGCAGCGGCTTGGTCTA 7560
Db 7501 CCATGCCCCCTTGGAGGGAGCGCTGGGATCCCGAATTCAGCAGCGGCTCATGCTGA 7560

QY 7561 CCCT---GAGTGAGGAGGCTAGTGAGGATCTCTGCTCTCAATGCTCTTACGTGA 7617
Db 7561 CGGTCACTAGTGGGGCGGACACCGAAGATGTCTGTCTCTCAATGCTCTTATACCTGGA 7620
QY 7618 CAGCGCCCTGATCAGCCCATGCGCTCGGAGGAAAGTAAAGCTGCCCATCAACCCCTTGA 7677
Db 7621 CAGCGCACTCGTCA CCGCGTGGCTGGGAGAAACAAAAACTGCCATCAACGCACTGA 7680
QY 7678 GCAACTCTTGTGCTGCTGCTACCAACAAATGCTTACGCGCAACATCCCGCAGCGCAAGCC 7737
Db 7681 GCAACTCTTGTGCTGCTGCTGCTACCAATCTGATATTTCCACCACTTCAACGAGTGTGC 7740
QY 7738 TCCGCGAGAGAAAGGTCACCTTTGACAGATTGCAAGTCTGCGATGATCATTAACCGGAGC 7797
Db 7741 AAAGGAGAGAAAGTCA CACTTTGACAGACTGCAAGTCTTGGACAGCACTTACAGGAGC 7800
QY 7798 TACTCAAGGATGAAGGAGAGCGCTCCACAGTTAAGGCTAAGCTTCTATCTATAGAGG 7857
Db 7801 TGTCTCAAGGAGGTCAAAGGAGCGCGCTCAAAAGTGAAGGCTAACTTGTCTATCCGTAGAGG 7860
QY 7858 AGGCTTGCAGGCTGACGCGCCCACTTTCGGCCAAATTCCTAAATTTGGCTATGGGGCAAGG 7917
Db 7861 AAGCTTGCAGCTGACGCGCCCACTTTCAGCCAAATTCCTAAATTTGGCTATGGGGCAAGG 7920
QY 7918 ACGTCCGGAACCTATCCAGCAGGGCGGTTAA CCACTCCGCTCCGCTGGGAGGACTTGC 7977
Db 7921 ACGTCCGTTGCCATGCCAGAAAGCGGTAGCCCACTCACTCCGTGTGGAAGGACTTTC 7980
QY 7978 TGAAGACACTGAACACCACTTGA CCAACCATATGCGCAAAAGTGAAGTTCGTCG 8037
Db 7981 TGAAGACAGTGTAA CCACTATAGACACTATCATATGCGCAAGGAGGCTTCTTCGCG 8040
QY 8038 TCCAAACAGAGAGGAGGCGCGAGCTCGCTTATCGCTTATTCCTCCAGACCTGGGAG 8097
Db 8041 TTCAGCTCAGAGGGGGTCTGTAAGCAGCTCTCATCTGTTTCCCGAGCTGGGCG 8100
QY 8098 TTCTGTATGCGAGAGATGGCCCTTTACAGCTGTCTTCCACCTTCTCTCAGGCGTGA 8157
Db 8101 TGGCGGTGTGCGAGAGATGGCCCTGTACAGCTGTGTAGCAAACTCCCTTGGCCGTGA 8160
QY 8158 TGGGCTCTCATACCGTTTCAATCTATGCTATGACACCGCTGTTTGTGACTCA 8277
Db 8161 TGGGAGCTCTTACCGATTCCAATCTCAGAGGACAGCGGTTGAAATTCCTCTGCTCAAG 8220
QY 8218 CCGTGAATCAAGAAATTCCTATGCGCTTATCGCGCTGCGCGCAAGTGGCGTGTGACGA 8457
Db 8221 CGTGAAGTCCAGAGAACCCCGATGGGGTCCCGTATGATACCCGCTGTTGACTCA 8280
QY 8278 CGGTCACTGAGGTGACATTTGTTGAGGAGTCAATTTACCAATGTTGTGACTTGGGCC 8337
Db 8281 CAGTCACTGAGAGGACATCCGTA CCGGAGGAGGCAATTTACCAATGTTGTGACTGGACC 8340
QY 8338 CCGAGCCAGACAGGCCATAGCTGCTTACAGAGCGGCTTATCATCGGGGTTCCCTGA 8397
Db 8341 CCCAGCCGCTGCGCAATCAAGTCCCTCACTGAGAGGCTTATGTTGGGGGCGCTCTTA 8400
QY 8398 CTAACTCAAAAGGAGAACTCGGCTTATCGCGCTGCGCGCAAGTGGCGTGTGACGA 8457
Db 8401 CCAATTCAGGGGGGAAACTGCGCTATCGAGGTGCGCGCGAGCGGCTACTGACAA 8460
QY 8458 CTAGTGGGTAAATACCTCTCATGTTTCTTGAAGGCACTGCAAGCTGTGAGCTGCAA 8517
Db 8461 CTAGTGTGTAAACCTCTCACTTGTACATCAAGGCGCGGAGCGGCTGAGCCGAG 8520
QY 8518 AGCTCCAGGACTGACGATGCTGTAACGAGAGGAGCTTGTGTTATCTGTGAAAGCG 8577
Db 8521 GGCTCCAGGACTGACCATGCTCTGTGTGGCGGAGCTTAGTCTGTTATCTGTGAAAGTG 8580
QY 8578 CCGGAAACCCAGAGGATGCGGCGGCTTACAGGCTTTCAGGAGGCTATGACTAGGTATT 8637
Db 8581 CCGGGGCTCAGAGGAGCGCGGAGGCTGAGGCTTTACGAGGCTATGACGAGGACT 8640
QY 8638 CCGCCCCCGGGGATCCGCCCCCAACAGAAATACGACCTGGAGCTGATAACATCATGTT 8697

Db 8641 CCGCCCCCCCCGGGACCCCGCCACACACAGAAATACGACTTGGAGCTTATAACATCATGCT 8700
QY 8698 CTTCAATGTGTCAFTCGGCGACGATGATCTGGCAAAAGGATATACTACCTCACCCGTG 8757
Db 8701 CTTCCAAACGTGTCAFTCGGCGACGATGATCTGGCAAAAGGATATACTACCTCACCCGTG 8760
QY 8758 ACCCCACACACCCCTTGCACGGCTGCTGGAGACAGCTAGACACACTCCAAATCAACT 8817
Db 8761 ACCCTACAAACCCCTTGCAGAGCGCTGGAGACAGCAAGACACACTCCAGTCAATT 8820
QY 8818 CTTGGCTAGGCAATATCATCTATGCGCCACCCCTATGGGCAAGGATGATTCGATGA 8877
Db 8821 CTTGGCTAGGCAATATCATCTATGCGCCACCCCTATGGGCAAGGATGATTCGATGA 8880
QY 8878 CTCACITTTTCTCCATCCTTCTAGTCTAAGACCAACTTGAATAAGCCCTGGATTCGAGA 8937
Db 8881 CCCATTTCTTAGCGTCTCATAGCCAGGATCAGCTTGAACAGGCTCTTAACCTGAGA 8940
QY 8938 TCTAGGGGCTTGTACTCCATTTGAGCCACTTGACCTACCTCAGATCATTTGAACACTCC 8997
Db 8941 TCTAGCGAGCTTGTACTCCATTTGAGCCACTTGAGCAACCACTGGATCTACCTCCATCATTTCAAAGACTCC 9000
QY 8998 ATGHTCTTAGCGCATTTTACCTCCACAGTTACTCTCCAGGTGAGATCAATAGGTTGGCTT 9057
Db 9001 ATGGCTCAGCGCATTTTACTCCACAGTTACTCTCCAGGTGAGTCAATAGGTTGGCTT 9060
QY 9058 CATGCTCAGAAACTTGGGTTACCAACCTTGGAAACCTGGAGACATCGGGCCAGAGTG 9117
Db 9061 CATGCTCAGAAACTTGGGTTACCAACCTTGGAAACCTGGAGACATCGGGCCAGAGTG 9120
QY 9118 TCCGGCTTAAGCTAGTCTCCAGGGGGAGGGCGGCACTGTGGGAGATACCTCTTTA 9177
Db 9121 TCCGGCTTAGCTTCTGTCCAGGGGAGGAGGGCTGCCATATGTGGCAAGTACCTCTTCA 9180
QY 9178 ACTGGGAGTAAAGACCAAGCTTAAACTCACTCCATCCCGGCGCGTCCAGCTGGACT 9237
Db 9181 ACTGGGAGTAAAGACCAAGCTTAAACTCACTCCATAGCGCGCGTGGCGGCTGGACT 9240
QY 9238 TGTCTGGCTGGTTGCTCGCTGGTTACAGCGGGGAGACATATACAGCCTGTCTGCTG 9297
Db 9241 TGTCTGGCTGGTTGCTCGCTGGTTACAGCGGGGAGACATATACAGCCTGTCTGCTG 9300
QY 9298 CCGACCCCGCTGTTTCCGTTGTCCTACTCTCTACTCTTCTGTAGGGGTAGGCAATTAAC 9357
Db 9301 CCGGCCCCCGCTGTTCTGTTTTCCTACTCTCTGCTGCTGCGTGGAGGGGTAGGCAATTAAC 9360
QY 9358 TGCTCCCAACCGATGAACGGGAGCTAACACCTCCAGGCTTAAGCAATTTCC 9411
Db 9361 TCCTCCCAACCGGTGAAGATTGGGCTAACCACTCCAGGCAATAGGCCATCCC 9414

Search completed: April 9, 2004, 08:49:33
Job time : 3079 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2004, 01:22:18 ; Search time 583 seconds
(without alignments)
9133.368 Million cell updates/sec

Title: US-09-662-454-4

Perfect score: 9595

Sequence: 1 gccagccccctgatgggggc.....gscctctctgcagatcatgt 9595

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9595	100.0	9595	3	US-09-014-416-4
2	8170.4	85.2	9416	1	US-08-324-977-1
3	8170.4	85.2	9416	2	US-08-384-616-1
4	8170.4	85.2	9416	2	US-08-904-686A-1
5	8170.4	85.2	9416	3	US-09-315-850-1
6	8170.4	85.2	9416	4	US-08-823-895A-27
7	8098.2	84.4	9472	4	US-08-150-204E-96
8	7918.6	82.5	11076	4	US-09-533-601-1
9	7901	82.3	11076	4	US-09-533-601-25
10	7899.4	82.3	11076	4	US-09-533-601-19
11	7894.6	82.3	11076	4	US-09-533-601-31
12	7809.2	81.4	9030	1	US-08-324-977-13
13	7809.2	81.4	9030	2	US-08-384-616-13
14	7809.2	81.4	9030	2	US-08-904-686A-13
15	7809.2	81.4	9030	3	US-09-315-850-13
16	7201.4	75.1	9599	3	US-09-014-416-6
17	6791.2	70.8	7917	1	US-08-324-977-31
18	6791.2	70.8	7917	2	US-08-384-616-31
19	6791.2	70.8	7917	2	US-08-904-686A-31
20	6791.2	70.8	7917	3	US-09-315-850-31
21	6754.8	70.4	7863	1	US-08-324-977-35
22	6754.8	70.4	7863	2	US-08-384-616-35
23	6754.8	70.4	7863	2	US-08-904-686A-35
24	6754.8	70.4	7863	3	US-09-315-850-35
25	6335.8	56.0	9599	3	US-09-014-416-2
26	6264	55.3	12980	3	US-08-811-566-5
27	6264	55.3	12980	4	US-09-034-756-5

28 6254.8 65.2 9646 3 US-08-811-566-1 Sequence 1, Appli
29 6254.8 65.2 9646 4 US-09-034-756-1 Sequence 1, Appli
30 6168.4 64.3 9401 5 PCT-US91-02225-9 Sequence 9, Appli
31 6166.4 64.3 9401 1 US-07-910-760-9 Sequence 9, Appli
32 6166.4 64.3 9401 1 US-08-440-519-9 Sequence 9, Appli
33 6166.4 64.3 9401 4 US-08-440-549-9 Sequence 9, Appli
34 6166.4 64.3 9401 4 US-08-823-895A-25 Sequence 25, Appli
35 6145.6 64.1 9379 3 US-08-444-818-176 Sequence 176, Appli
36 6144.4 64.0 9379 3 US-09-388-874-1 Sequence 1, Appli
37 6135.6 63.9 9416 3 US-09-916-359-1 Sequence 1, Appli
38 6135.6 63.9 9416 3 US-08-811-566-19 Sequence 19, Appli
39 6134.4 63.9 9401 2 US-09-034-756-19 Sequence 19, Appli
40 6134.4 63.9 9401 2 US-08-432-693-1 Sequence 1, Appli
41 6113.2 63.7 9416 4 US-08-823-895A-26 Sequence 26, Appli
42 6113.2 63.7 9416 4 US-10-104-966-13 Sequence 13, Appli
43 5999.4 62.5 9185 3 US-08-444-818-122 Sequence 122, App
44 5999.4 62.5 9185 3 US-08-444-818-123 Sequence 123, App
45 5854 61.0 8637 4 US-09-539-601-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-014-416-4
; Sequence 4, Application US/09014416
; Patent No. 6153421

; GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Buhi, Jens

; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.

; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416

; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062

; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4

; LENGTH: 9595
; TYPE: DNA

; ORGANISM: Hepatitis C virus
US-09-014-416-4

09/662,454
Examiner's
Search Notes
5/17/04

Query Match 100.0%; Score 9595; DB 3; Length 9595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9595; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCAGCCCCCTGATGGGGGCGACACTCCACATGATCACTCCCTGTGAGGAACACTAGT 60
Db 1 GCCAGCCCCCTGATGGGGGCGACACTCCACATGATCACTCCCTGTGAGGAACACTAGT 60

QY 61 TCTTCAGCGAGAAGCGCTCTAGCCATGCGCTGTAGTATGCTGTCGACACCTCCAGAC 120
Db 61 TCTTCAGCGAGAAGCGCTCTAGCCATGCGCTGTAGTATGCTGTCGACACCTCCAGAC 120

QY 121 CCCCCCTCCCGGAGAGCCACTAGTGGTCTCGGAAACCGGTGAGTACACCGGAATTGCCAG 180
Db 121 CCCCCCTCCCGGAGAGCCACTAGTGGTCTCGGAAACCGGTGAGTACACCGGAATTGCCAG 180

QY 181 GACGACCGGTCTCTTTCTTGGATCAACCCCTCAATGCCCTGGAGATTGGGCGTGCCTCC 240
Db 181 GACGACCGGTCTCTTTCTTGGATCAACCCCTCAATGCCCTGGAGATTGGGCGTGCCTCC 240

QY 241 GCGAGACTGTAGCCGAGTAGTGTGGTTCGGAAGCCCTTGTGTACTGCTGATAGG 300
Db 241 GCGAGACTGTAGCCGAGTAGTGTGGTTCGGAAGCCCTTGTGTACTGCTGATAGG 300

QY 301 GTGCTTCGAGTAGTCCCGGAGGTCTGCTAGACCGTGCACCATGAGCAGCAATCTTAAC 360
Db 301 GTGCTTCGAGTAGTCCCGGAGGTCTGCTAGACCGTGCACCATGAGCAGCAATCTTAAC 360

Db 301 GTGCTTGGCGAGTGCCCCGGGAGGHTCTGTAGACCGTGCACCAATGACGACGATCCTAAAC 360
Qy 361 CTCAAAGAAACAAACCTTAACACAAACCGCGCGCCACAGACGCTCAAGTTCCCGGGG 420
Db 361 CTCAAAGAAACAAACCTTAACACAAACCGCGCGCCACAGACGCTCAAGTTCCCGGGG 420
Qy 421 GTGGTCAGATGTTGGTGAGTTTACCTGTTGCCGCGAGGGGCCAGGTTGGGTGTGC 480
Db 421 GTGGTCAGATGTTGGTGAGTTTACCTGTTGCCGCGAGGGGCCAGGTTGGGTGTGC 480
Qy 481 GCCCGACTAGGAAGGCTTCCGAGCGGTGCGCAACCTCTGTGGAGGCGCAACCTATCCAA 540
Db 481 GCGCGACTAGGAAGGCTTCCGAGCGGTGCGCAACCTCTGTGGAGGCGCAACCTATCCAA 540
Qy 541 AGGCTTCGCGACCCGAGGCGAGGCTTGGGCTCAGCCCGGGTACCCCTTGGCCCTCTATG 600
Db 541 AGGCTTCGCGACCCGAGGCGAGGCTTGGGCTCAGCCCGGGTACCCCTTGGCCCTCTATG 600
Qy 601 GCAATGAGGCTTGGGTTGGGAGGATGGCTCTGTCAACCGCGGCTCCCGGCTAGTT 660
Db 601 GCAATGAGGCTTGGGTTGGGAGGATGGCTCTGTCAACCGCGGCTCCCGGCTAGTT 660
Qy 661 GGGCCCCACGACCCCGCGGTAGGTCGCTTAACCTTGGGTAAGGTCACTGATACCTTA 720
Db 661 GGGCCCCACGACCCCGCGGTAGGTCGCTTAACCTTGGGTAAGGTCACTGATACCTTA 720
Qy 721 CATGGGCTTCGCGATCTCATGGGTACATTCGCTCTGTCGGCGCCCGCTAGGGGCG 780
Db 721 CATGGGCTTCGCGATCTCATGGGTACATTCGCTCTGTCGGCGCCCGCTAGGGGCG 780
Qy 781 CTGCCAGGCTTCGCGACCGGTCGCGGTTCGGAGGACGCGGTGAACATGCAACAG 840
Db 781 CTGCCAGGCTTCGCGACCGGTCGCGGTTCGGAGGACGCGGTGAACATGCAACAG 840
Qy 841 GGAACCTGCGCGTGTCTTCTATCTTCTCTATCTTCTCTGCTGCTGCTGCTGTGACCA 900
Db 841 GGAACCTGCGCGTGTCTTCTATCTTCTCTATCTTCTCTGCTGCTGCTGCTGTGACCA 900
Qy 901 TCCAGCTTCGCGTATGAAGTGGCAACGTTCCGGGATATACCATGTCAACGACGACT 960
Db 901 TCCAGCTTCGCGTATGAAGTGGCAACGTTCCGGGATATACCATGTCAACGACGACT 960
Qy 961 GCTCCAACTCAAGCATGTGATAGGACGCGGAGGTGATCATGATCTCCGGGTGG 1020
Db 961 GCTCCAACTCAAGCATGTGATAGGACGCGGAGGTGATCATGATCTCCGGGTGG 1020
Qy 1021 TGCCTGTGTTTCAGAGGGTAAACAGTCCCGTGTCTGGTACGCTCACTCCGAGCTCG 1080
Db 1021 TGCCTGTGTTTCAGAGGGTAAACAGTCCCGTGTCTGGTACGCTCACTCCGAGCTCG 1080
Qy 1081 CGGCGAGGATGCCAGGTCCCACTAGCAATACGACGCAAGTGTGCTCGTTG 1140
Db 1081 CGGCGAGGATGCCAGGTCCCACTAGCAATACGACGCAAGTGTGCTCGTTG 1140
Qy 1141 GGAAGGCTGCTTCTGCTCCGCTATGATGAGGAGTCTCTGGGATCTCTGCGATCTATTTCTCG 1200
Db 1141 GGAAGGCTGCTTCTGCTCCGCTATGATGAGGAGTCTCTGGGATCTCTGCGATCTATTTCTCG 1200
Qy 1201 TCTCCGAGCTTTCACCTTCTGCTCCGCGCATGAGACGAGTCCAGGACTGCAACTGCT 1260
Db 1201 TCTCCGAGCTTTCACCTTCTGCTCCGCGCATGAGACGAGTCCAGGACTGCAACTGCT 1260
Qy 1261 CAATCTATCCGGCGCATGTATCAGGTCAACGCTGGCTTGGATATGATGAACTGTT 1320
Db 1261 CAATCTATCCGGCGCATGTATCAGGTCAACGCTGGCTTGGATATGATGAACTGTT 1320
Qy 1321 CACCTACACAGCCCTAGTGTGTCGAGTTGCTCCGATGCCACACAGCTGTCTGAGCA 1380
Db 1321 CACCTACACAGCCCTAGTGTGTCGAGTTGCTCCGATGCCACACAGCTGTCTGAGCA 1380
Qy 1381 TGTGGCGGGGCCCACTGGGAGTCTCTGGCGGCTTGGCTACTATTCATGTTAGGGA 1440
Db 1381 TGTGGCGGGGCCCACTGGGAGTCTCTGGCGGCTTGGCTACTATTCATGTTAGGGA 1440

Qy 1441 ACTGGCTAGGTTCTGATTTGTGGGCTACTCTTTGGCGGCTTGACGGGAGAGCCACACA 1500
Db 1441 ACTGGCTAGGTTCTGATTTGTGGGCTACTCTTTGGCGGCTTGACGGGAGAGCCACACA 1500
Qy 1501 CGACGGGAGGGTGGCGGCGCACACCACTCCGGGTTCAAGTCCCTTTCTCATCTGGGG 1560
Db 1501 CGACGGGAGGGTGGCGGCGCACACCACTCCGGGTTCAAGTCCCTTTCTCATCTGGGG 1560
Qy 1561 GGTCTCAGAAATCCAGCTTGTGAATACCAACGCGACGCTGGCAGCATCAACAGGACTGCC 1620
Db 1561 GGTCTCAGAAATCCAGCTTGTGAATACCAACGCGACGCTGGCAGCATCAACAGGACTGCC 1620
Qy 1621 TAAATTCGAATGACTCCCTCAAACTGGGTTCTTTGGCGGCTGTTTACGCAACAGT 1680
Db 1621 TAAATTCGAATGACTCCCTCAAACTGGGTTCTTTGGCGGCTGTTTACGCAACAGT 1680
Qy 1681 TCAACTCGTCGGGTGCCCGAGCGCATGCGCAGCTCCCGCCCAATGACTGGTTCCGCC 1740
Db 1681 TCAACTCGTCGGGTGCCCGAGCGCATGCGCAGCTCCCGCCCAATGACTGGTTCCGCC 1740
Qy 1741 AGGGTGGGCGCCATCACTATTAAGCTTAACAGCTCGATCGATCGAGGCTTTATGCT 1800
Db 1741 AGGGTGGGCGCCATCACTATTAAGCTTAACAGCTCGATCGATCGAGGCTTTATGCT 1800
Qy 1801 GGCATTAACGCGCTCGACCGCTGTGCTACCCGCGTCCGAGGTGTGTCCAGTGT 1860
Db 1801 GGCATTAACGCGCTCGACCGCTGTGCTACCCGCGTCCGAGGTGTGTCCAGTGT 1860
Qy 1861 ATTGTTTCAACCCAAAGCCCTGTTGTTGGTGGGACCAACCGATCGTTCCGTTCCCTACGT 1920
Db 1861 ATTGTTTCAACCCAAAGCCCTGTTGTTGGTGGGACCAACCGATCGTTCCGTTCCCTACGT 1920
Qy 1921 ATAGCTGGGAGGATGAGACAGCTGATCTCTCAACACACGCTCCGGCAACAG 1980
Db 1921 ATAGCTGGGAGGATGAGACAGCTGATCTCTCAACACACGCTCCGGCAACAG 1980
Qy 1981 GCAACTGTTTCGGTGTACATGATGAATAGTACTTGGTTTCACTTAAGCTGCGGAGSTC 2040
Db 1981 GCAACTGTTTCGGTGTACATGATGAATAGTACTTGGTTTCACTTAAGCTGCGGAGSTC 2040
Qy 2041 CCCCGTGTAAACATCGGGGGTTCGTTAACCGCACTTGATCTGCCCCACGCACTGCTTC 2100
Db 2041 CCCCGTGTAAACATCGGGGGTTCGTTAACCGCACTTGATCTGCCCCACGCACTGCTTC 2100
Qy 2101 GGAAGCACCCCGAGGCTTACACAAATGTGGCTGGGGCCCTGGTGAACCTAGGT 2160
Db 2101 GGAAGCACCCCGAGGCTTACACAAATGTGGCTGGGGCCCTGGTGAACCTAGGT 2160
Qy 2161 GCCTAGTAGACTACCCATACAGGCTTGGCACTTACCCCTGCACTCTCAATTTTCCATCT 2220
Db 2161 GCCTAGTAGACTACCCATACAGGCTTGGCACTTACCCCTGCACTCTCAATTTTCCATCT 2220
Qy 2221 TTAAGTTAGGATGTATGTGGGGGCGTGAGCACAGGCTCAATGCGCATGCAATGGA 2280
Db 2221 TTAAGTTAGGATGTATGTGGGGGCGTGAGCACAGGCTCAATGCGCATGCAATGGA 2280
Qy 2281 CTCGAGAGAGCTGTAACTTGGAGACAGGATAGTCAAGATCAGCCGCTCTGC 2340
Db 2281 CTCGAGAGAGCTGTAACTTGGAGACAGGATAGTCAAGATCAGCCGCTCTGC 2340
Qy 2341 TGTCTCAACAGAGTGGCAGATCTCCCTGTGCTTTCACCACTTACCGGCTTTATCCA 2400
Db 2341 TGTCTCAACAGAGTGGCAGATCTCCCTGTGCTTTCACCACTTACCGGCTTTATCCA 2400
Qy 2401 CTGGTTTGTATCCATCTCCATACAGACATCGTGAAGTCAATACCTGTACGTTAGGCT 2460
Db 2401 CTGGTTTGTATCCATCTCCATACAGACATCGTGAAGTCAATACCTGTACGTTAGGCT 2460
Qy 2461 CAGCGTTCTCTCCCTTGCATTCGAATGGAGGTACATCCTGTTGCTTTTCTTCTCTCG 2520
Db 2461 CAGCGTTCTCTCCCTTGCATTCGAATGGAGGTACATCCTGTTGCTTTTCTTCTCTCG 2520

Db	4691	ATTTTGACTCAGTCAGTCAGTCGCAATACATGTTCTACCCAGACAGTGCAGTTCAGCTTGC	4741
Qy	4741	ATCCACCTTACCATTACAGACGACGACCGTGTCCCCAAAGACGCGGTGTGCGCTCGCAAC	4800
Db	4741	ATCCACCTTACCATTACAGACGACGACCGTGTCCCCAAAGACGCGGTGTGCGCTCGCAAC	4800
Qy	4801	GGCAGGTAGAACTGGCAGGGGTAGGAGTGGCATCTACAGTTTGTGACTCCAGGAGAAC	4860
Db	4801	GGCAGGTAGAACTGGCAGGGGTAGGAGTGGCATCTACAGTTTGTGACTCCAGGAGAAC	4860
Qy	4861	GGCCCTCGGGCATGTTGCATTTCTCGGTCGTCGTGAGTGCATGACGCGGGGTGTCCTT	4920
Db	4861	GGCCCTCGGGCATGTTGCATTTCTCGGTCGTCGTGAGTGCATGACGCGGGGTGTCCTT	4920
Qy	4921	GGTATGAGCTCACGCCCGCTGAGACCTCGGTAGGTTGCGGGCTTACTTAATACACAG	4980
Db	4921	GGTATGAGCTCACGCCCGCTGAGACCTCGGTAGGTTGCGGGCTTACTTAATACACAG	4980
Qy	4981	GGTTGCCCGTCTGCCAGGACCAATCTGGAGTTCTGGGAGAGGCTCTTCAAGGGCTCACCC	5040
Db	4981	GGTTGCCCGTCTGCCAGGACCAATCTGGAGTTCTGGGAGAGGCTCTTCAAGGGCTCACCC	5040
Qy	5041	ACATAGATGCCACTTCTGTCGCCAGACTAAACAGCAGCAGAGCAACTTTCTTCACTCG	5100
Db	5041	ACATAGATGCCACTTCTGTCGCCAGACTAAACAGCAGCAGAGCAACTTTCTTCACTCG	5100
Qy	5101	TGGCATATCAAGCTPACAGTGTGGCCAGGGCTCAAGCTCCACCTCCATCGTGGGACCAAA	5160
Db	5101	TGGCATATCAAGCTPACAGTGTGGCCAGGGCTCAAGCTCCACCTCCATCGTGGGACCAAA	5160
Qy	5161	TGTGGAGTGTCTCATAGGCTGAACCTACATGCA CGGGCCAAACACCCTGCTGTPATA	5220
Db	5161	TGTGGAGTGTCTCATAGGCTGAACCTACATGCA CGGGCCAAACACCCTGCTGTPATA	5220
Qy	5221	GGCTAGCAGCGCTCCAAAATGAGTGCATCTTCA CACACCCCAATCAATCATCATG	5280
Db	5221	GGCTAGCAGCGCTCCAAAATGAGTGCATCTTCA CACACCCCAATCAATCATCATG	5280
Qy	5281	CATGCATGTCGGCTGACTGGAGTGTCTACTAGCACTGGGTGCTGCTAGGCGGAGTCC	5340
Db	5281	CATGCATGTCGGCTGACTGGAGTGTCTACTAGCACTGGGTGCTGCTAGGCGGAGTCC	5340
Qy	5341	TTGCAGCTTTGGCGGCTACTGCTGACAGCAGCAGTGTGTCATTTGTCGGCAGGATCA	5400
Db	5341	TTGCAGCTTTGGCGGCTACTGCTGACAGCAGCAGTGTGTCATTTGTCGGCAGGATCA	5400
Qy	5401	TCTTGTCCGGGAAGCCAGCTGTGCTTCCGACAGGAGTCCCTCA CCGAGTTCGATG	5460
Db	5401	TCTTGTCCGGGAAGCCAGCTGTGCTTCCGACAGGAGTCCCTCA CCGAGTTCGATG	5460
Qy	5461	AGATGGAAGAGTGTGCTCTCAACTTCTTACATCGAGCAGGGAATGCAGCTCCCGAGC	5520
Db	5461	AGATGGAAGAGTGTGCTCTCAACTTCTTACATCGAGCAGGGAATGCAGCTCCCGAGC	5520
Qy	5521	AATTCAAGCAAAAGGCGCTCGGTTGTTGCAAGCGCCACCAAGCAAGCGAGGCTGCTG	5580
Db	5521	AATTCAAGCAAAAGGCGCTCGGTTGTTGCAAGCGCCACCAAGCAAGCGAGGCTGCTG	5580
Qy	5581	CTCCGCTGTGTGAGTCCAGTGGCGCCCTTCGAGACCTTCGGGCGAAGCATATGGA	5640
Db	5581	CTCCGCTGTGTGAGTCCAGTGGCGCCCTTCGAGACCTTCGGGCGAAGCATATGGA	5640
Qy	5641	ATTTTCATCAGCGGAATACAGTACTAGCAGGCTTATCCACTCTGCTGGAACCCCGCA	5700
Db	5641	ATTTTCATCAGCGGAATACAGTACTAGCAGGCTTATCCACTCTGCTGGAACCCCGCA	5700
Qy	5701	TAGCATCATTTGATGGCATTTTACAGTTCTATCTAGCCTGCTCA CCAACCAAAACACCC	5760
Db	5701	TAGCATCATTTGATGGCATTTTACAGTTCTATCTAGCCTGCTCA CCAACCAAAACACCC	5760
Qy	5761	TCTCTGTTTAACTCTTGGGGGATGGGTGGCTGCCCAACTCTGCTCTCCAGCGCTGCTG	5820
Db	5761	TCTCTGTTTAACTCTTGGGGGATGGGTGGCTGCCCAACTCTGCTCTCCAGCGCTGCTG	5820

Qy	5821	CAGCTTTCGTGGCGCCCGCATCCCGAGCGGCTGTGTGCACATAGGCGCTTTGGGAAGG	5880
Dz	5821	CAGCTTTTCGTGGCGCCCGGATCCCGGAGCGGCTGTGGCAGCATAGGCGCTTTGGGAAGG	5880
Qy	5881	TGCTCGTGGACATCTTCGCGGGCTATGGGCGAGGGGTAGCGGCGCACTCGTCGCGCTTTA	5940
Dz	5881	TGCTCGTGGACATCTTCGCGGGCTATGGGCGAGGGGTAGCGGCGCACTCGTGCGCTTTA	5940
Qy	5941	AGGTCATGAGCGCGAGGTCCCTCCACCGAGGACCTGTGTAACTTACTCCCTGCCATCC	6000
Dz	5941	AGGTCATGAGCGCGAGGTCCCTCCACCGAGGACCTGTGTAACTTACTCCCTGCCATCC	6000
Qy	6001	TCTCTCTGTGTCCCTCGTCCGGGTCTGTGTCCGCGCAATACTTGGTCGCGCTGG	6060
Dz	6001	TCTCTCTGTGTCCCTCGTCCGGGTCTGTGTCCGCGCAATACTTGGTCGCGCACTGG	6060
Qy	6061	GCCCGGAGAGGGGCTGTGTGAGTGAATAACCGGCTGTAGCTTTCGTCGCGGGTA	6120
Dz	6061	GCCCGGAGAGGGGCTGTGTGAGTGAATAACCGGCTGTAGCTTTCGTCGCGGGTA	6120
Qy	6121	ACCACTGTCCCTACGCACTATGTGCTGAGAGGACGCTGCAGACGCTGTCACTCAGA	6180
Dz	6121	ACCACTGTCCCTACGCACTATGTGCTGAGAGGACGCTGCAGACGCTGTCACTCAGA	6180
Qy	6181	TCCCTCTAGCTTTACCATCACTCAACTGTGAAGCGGCTCCACCACTGGATTAATGAGG	6240
Dz	6181	TCCCTCTAGCTTTACCATCACTCAACTGTGAAGCGGCTCCACCACTGGATTAATGAGG	6240
Qy	6241	ACTGCTCTAGCANTGTCCGGCTCGTGTGCTAAGGATGTTTGGATTTGATGACCGG	6300
Dz	6241	ACTGCTCTAGCANTGTCCGGCTCGTGTGCTAAGGATGTTTGGATTTGATGACCGG	6300
Qy	6301	TGTTGACTGACTTCAAGACCTGCTCCAGTCCAAACTCTCGCGCGGTTACCGGAGTCC	6360
Dz	6301	TGTTGACTGACTTCAAGACCTGCTCCAGTCCAAACTCTCGCGCGGTTACCGGAGTCC	6360
Qy	6361	CTTTCCTGTCTATGCCAACCGGGTACAAGGAGTCTGGCGGGGAGCGGCATCATGCAAA	6420
Dz	6361	CTTTCCTGTCTATGCCAACCGGGTACAAGGAGTCTGGCGGGGAGCGGCATCATGCAAA	6420
Qy	6421	CCACTGCCATGGGAGCACAGATCCGGGACATGTCAAAACGTTCCATGTGGATCG	6480
Dz	6421	CCACTGCCATGGGAGCACAGATCCGGGACATGTCAAAACGTTCCATGTGGATCG	6480
Qy	6481	TAGGGCTTAGAAGCTGCAGCAACGCTGGCACGGAACTTCCCATCAACGATACACCA	6540
Dz	6481	TAGGGCTTAGAAGCTGCAGCAACGCTGGCACGGAACTTCCCATCAACGATACACCA	6540
Qy	6541	CGGACCTTGACACACCTCCCGGCGCCCACTATTCCAGGGGCTATGCGGGTGGCTG	6600
Dz	6541	CGGACCTTGACACACCTCCCGGCGCCCACTATTCCAGGGGCTATGCGGGTGGCTG	6600
Qy	6601	CTGAGGAGTACGTGGAGGTATCCGCTGTGGGGATTTCCACTACGTACGGCATGACCA	6660
Dz	6601	CTGAGGAGTACGTGGAGGTATCCGCTGTGGGGATTTCCACTACGTACGGCATGACCA	6660
Qy	6661	CTGACACGTAAAGTCCCATGCGAGTTCCGGGCCCGAAATTCCTTACGGAGTGGATG	6720
Dz	6661	CTGACACGTAAAGTCCCATGCGAGTTCCGGGCCCGCGAAATTCCTTACGGAGTGGATG	6720
Qy	6721	GAGTGGCGTTGACAGGTACGCTCGGCGGTGAAACCTCTTCTACGGAGGACGTACGT	6780
Dz	6721	GAGTGGCGTTGACAGGTACGCTCGGCGGTGAAACCTCTTCTACGGAGGACGTACGT	6780
Qy	6781	TCCAGTTCGGGCTCAACCAATATTGGTTCGGGTCGAGCTCCATGCGAGCCCGAACCGG	6840
Dz	6781	TCCAGTTCGGGCTCAACCAATATTGGTTCGGGTCGAGCTCCATGCGAGCCCGAACCGG	6840
Qy	6841	ACCTAACGTGCTTACTTCCATGTCTCACGATCCCTCCCACTATTACAGCAGACGGCTA	6900
Dz	6841	ACCTAACGTGCTTACTTCCATGTCTCACGATCCCTCCCACTATTACAGCAGACGGCTA	6900

QY 614 GGGGTGGGAGGATGGCTCTCTGATCCCGCGGCTCCCGGCTAGTTGGGGCCCGACGGA 673
 Db 605 AGGTGGGAGAGATGGCTCTCTGATCCCGCGGCTCCCGGCTAGTTGGGGCCCGACGGA 664
 QY 674 CCCCCGGGTAGGTGGCGTAATTTGGTAAAGTCAATCGATACCGTTCATATGGGTCTGC 733
 Db 665 CCCCCGGGTAGGTGGCGTAATTTGGTAAAGTCAATCGATACCGTTCATATGGGTCTGC 724
 QY 734 CGATCTCATGGGGTACATTCGCGCTCGTGGCGCCCGCTAGGGGGCGTCCAGGGCCCTT 793
 Db 725 CGATCTCATGGGGTACATTCGCGCTCGTGGCGCCCGCTAGGGGGCGTCCAGGGCCCTT 784
 QY 794 GGCACAGGTGTCGGGGTTCGGAGACGCGCTGAATATGCAACAGGGAATCTTCCCGG 853
 Db 785 GGCACATGTTGTCGGGGTTCGGAGACGCGCTGAATATGCAACAGGGAATCTTCCCGG 844
 QY 854 TTGCTCTTTCTCTATCT 913
 Db 845 TTGCTCTTTCTCTATCT 904
 QY 914 TTATGAAGTGGCAACGCTGTCGGGATATACCATGTCACGACGACTGCTCCAACTCAAG 973
 Db 905 TTACGAAGTGGCAACGCTGTCGGGATATATCATGTCACGACGACTGCTCCAACTCAAG 964
 QY 974 CATTTGTATGAGGAGCGGAGCTGATCATGATATCTCCCGGCTGGTCCCTGTGTCTCA 1033
 Db 965 CATTTGTATGAGGAGCGGAGCTTGTATCATGATATCTCTGGTGGTGGCTGGCTTCCG 1024
 QY 1034 GGAGGTATACGCTCCGTTGCTGGGTAGCTGCTCACTCCAGCTCGCGGCGAGGAATGC 1093
 Db 1025 GGAAGGCAACTCTCCCGCTGCTGGGTAGCTGCTCACTCCAGCTCGCGGCGAGGAAGCT 1084
 QY 1094 CAGCTCTCCCACTACGACAAATACGACGCGCACGCTGCACTGCTGTTGGGACGCGTCTT 1153
 Db 1085 CACCAATCCCAACACGACGATACGACGCGCACGCTGATCTGCTGGTGGGCGGCTGCTT 1144
 QY 1154 CTGCTCCGCTATGATCGTGGGGATCTCTCGGATCTATTTCCCTCGTCTCCAGCTGTT 1213
 Db 1145 CTGTTCCGCTATGATCGTGGGGAACTCTGCGGATCTGTTTCTCTGCTCTCACTGTT 1204
 QY 1214 CACCTCTCGCTCGCGGCTAGACGATGAGAGTGCAGATGCAATCTGCTCAATCTACCGG 1273
 Db 1205 CACCTCTCGCTCGCGGCTAGTACATTAAGGACTGTAATCTGCTCAATTTATCCCGG 1264
 QY 1274 CCATGATCAGCTCACCGATCGCTGGGATGATGATGATGATGATGATGATGATGATGAT 1333
 Db 1265 CCACTGTCGAGTCAACGATAGCTTGGGACATGATGATGATGATGATGATGATGATGAT 1324
 QY 1334 CCTAGTGTGTCGAGTGTCTCCGATCCCAAGCTGTCGTCGACATGCTGGTGGCGGGGC 1393
 Db 1325 CCTAGTGTGTCGAGTGTCTCCGATCCCAAGCTGTCGTCGACATGCTGGTGGCGGGGC 1384
 QY 1394 CCACCTGGGAGTCTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1453
 Db 1385 CCACCTGGGAGTCTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1444
 QY 1454 TCTGATTTGGGGTACTCTTTGGCGGCTTGGAGGAGACCCACAGCGGAGAGGT 1513
 Db 1445 TCTGATTTGATGCTACTTTTGGCTGGGCTTGGAGGAGATACCCAGTGCACAGGGGGGC 1504
 QY 1514 GCGCGGCACACCACTCGGGTTCAGTCCCTTTCTCATCTGGGGCTCTCAGAAAT 1573
 Db 1505 GCAAGCCAAAACCAACAGGCTGCTGCTCCATGTTGGCAAGTGGGCGCTCTCAGAAAT 1564
 QY 1574 CCAGCTTGTATACCAAGGAGCTGGGACATCAACAGGACTGCCCTTAATGCAATGA 1633
 Db 1565 CCAGCTTGTATACCAAGGAGTGGGACATCAACAGGACTGCCCTTGAATGCAATGA 1624
 QY 1634 CTCCCTCAAACTGGGTCTTTGGCGGCTGTTTACGACACAAAGTTCAACTCTGTCGG 1693
 Db 1625 CTCTCTCAGACTGGGTCTTTGGCGGCTGTTTACACAAATGTTTCAACTCTGTCGG 1684
 QY 1694 GTGCCGAGAGCGCATGGGACGCTGCCGCCCAATGACTGTTTGGGCTGGGGGCTG 1753

Db 1685 GTGCCGAGCGCATGGCCCGAGTGGCGCACCATTTGACAGTTTGACACAGGATGGGGTCC 1744
 QY 1754 CATCACTTATATTAAGCCTAAACAGTTCGATACAGAGGCTTATTTGCTGGCAATTACGCGCC 1813
 Db 1745 CATTACTTATGCTGAGTCTAGCAGATCAGACAGAGGCGCATTTGCTGGCACTACCCACC 1804
 QY 1814 TCGACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1873
 Db 1805 TCCCAATGTTACCATTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1864
 QY 1874 AAGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1933
 Db 1865 AAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1924
 QY 1934 GAATGAGACAGACGCTGATGCTCTCTACACACGCGTCCGCCACAAGGCAACTGCTGCTGG 1993
 Db 1925 GAACGAGACTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1984
 QY 1994 CTGTACATGATGAATAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2053
 Db 1985 CTGCACATGATGAATAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2044
 QY 2054 CGGGGGGCTCGGTAAACCGCACCTTTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2113
 Db 2045 CGGGGGGCTCGGCAACCAACCTCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2104
 QY 2114 GGCTACTTACACAAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2173
 Db 2105 GGCTACTTACACAAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2164
 QY 2174 CCCATACAGGCTTTGGGCACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2233
 Db 2165 TCCATACAGGCTTTGGGCACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2224
 QY 2234 GTATGTTGGGGGCTGGGACACAGGCTCAATGCGCGCATGCAATTTGGACTCGAGGAGAGCG 2293
 Db 2225 GTATGTTGGGGGCTGGGACACAGGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2284
 QY 2294 CTGTAACTTTGAGGACAGGATAGGTGAGAACTCAGCCCGCTGCTGCTGCTGCTGCTGCTGCT 2353
 Db 2285 TTGTGACTTTGAGGACAGGATAGGCGCGGAGCTCAGCCCGCTGCTGCTGCTGCTGCTGCTGCT 2344
 QY 2354 GTGCGCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2413
 Db 2345 GTGCGCAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2404
 QY 2414 TCTCCATCAGAACTTCGTGGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2473
 Db 2405 CTTCCATCAGAACTTCGTGGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2464
 QY 2474 CTTTGCATCAAAATGGGAGTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2533
 Db 2465 CTTTGCATCAAAATGGGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2524
 QY 2534 GTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2593
 Db 2525 CTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2584
 QY 2594 GGTGCTCTCAATCGCGGCTCCGTGGCGGAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2653
 Db 2585 GGTGCTCTCAATCGCGGCTCTGTGGCGGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2644
 QY 2654 CTTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2713
 Db 2645 CTTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2704
 QY 2714 TGGGCTATGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2773
 Db 2705 TGGGCTATGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2764
 QY 2774 CCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2833

Qy	4994	CCAGGACCAATCTGGAGTTCTGGGAGAGGGCTCTTACAGAGGCTCTACCCACATAGATGCCCCA	5053
Db	4985	CCAGGACCACTGGAGTTCTGGGAGAGTGTCTTTCACAGGCTCCACCCATATAGATGACACA	5044
Qy	5054	CTTCTCTCCAGAGCTAAACAGGACAGAGACAACTTTTCCTTACCTGCTGGCATATCAAGC	5113
Db	5045	CTTCTTGTCCAGACCAAGCAGGACAGAGACAATCTCCCTTACCTGTGTAGCATACCAAGC	5104
Qy	5114	TACAGTGTGCOCACGGGCTCAAGCTCCACCTTCCATCTGTGGGACCAAAATGTGGAGTGTCT	5173
Db	5105	CACGGTGTGCOCAGGGCTCAGGCCCCACCTTCATCTATGGCATCAAAATGTGGAATGTCT	5164
Qy	5174	CATAGGGCTGAAACCTACACTGACAGGGGCAACACCCCTGCTGTATATAGGCTAGAGACCGT	5233
Db	5165	CATACGGCTGAAACCTACCGCTGCACGGGCAACACCTTGTCTGTACAGGCTGGGAGCGCT	5224
Qy	5234	CCAAAATGAGGTCTATCTCTCACACACCCCAATAAATACATCATCATGSCATGCAATCTCGC	5293
Db	5225	CCGAAATGAGGTCACTCCCTCACCCACCCCAATACATCATGSCATGCAATCTGCTCGC	5284
Qy	5294	TGAOCTGAGGTCTGCTCACTAGCACTGGGTCTGTGTAGCGGAGTCTCTTGACGCTTTTGGC	5353
Db	5285	TGAOCTGAGGTCTGCTCACTAGCACTGGGTCTGTGTAGCGGAGTCTCTTGACGCTCTTGGC	5344
Qy	5354	CGCATACTGCTTGAGGACAGGAGTGTGTCTATTTGTGTGGCAGGATCATCTTGTCTCGGAAA	5413
Db	5345	CGCGTATTGCTGTGACACAGGACAGTGTGTCTATTTGTGTGGTAGGATTAATCTTGTCTCGGAG	5404
Qy	5414	GCCAGCTGTCTCTCCGACAGGGAAGTCTCTTACAGAGGTTCCATGAGATGGAGAGTG	5473
Db	5405	GCCGGCCATGTCTTCCGACAGGAGCTTCTCTTACAGAGTTCGATGAATGGGAAGAGTG	5464
Qy	5474	TGCCTCACAACTTCTTTACATCGACGAGGGAATCAGCTCCCGACGAAATTCAGACAAA	5533
Db	5465	CGCTCTGCACCTCTCCTTATCATCGACGAGGGAATCAGCTCCCGACGAAATTCAGACAAA	5524
Qy	5534	GGCGCTGGGTGTGTGAAACGGGCAACAGCAAGCGAGGCTGCTGCTCCGCTGGTGGGA	5593
Db	5525	AGCGCTGGGTGTGTGAAACAGGCAACAGCAAGCGAGGCTGCTGCTCCGCTGGTGGGA	5584
Qy	5594	GTCCAAGTGGGAGCGCTTGAGACCTTCTGGCGGACGACATGTTGGAAATTCATCAGCGG	5653
Db	5585	GTCCAAGTGGGAGCGCTTGAGACCTTCTGGCGGACGACATGTTGGAAATTCATCAGCGG	5644
Qy	5654	AATACAGTACCTAGCAGGCTTATCCACTCTGCTCTGGAAACCCCGCGATAGCATCATTTGAT	5713
Db	5645	GATACAGTACTTACAGGCTTATCCACTCTGCTCTGGAAACCCCGCATAGCATCATTTGAT	5704
Qy	5714	GGCATTTACGCTTCTATCATAGCGGCTCACACCCAAACACCTCTCTGTTTAAACAT	5773
Db	5705	GGCATTTACAGCTCTATCACAGCGGCTCACACCCAAAGTACCCTCTCTGTTTAAACAT	5764
Qy	5774	CTTGGGGGATGGGTGGCTGCCAACTGCGTCTCCAGCGGCTCGGTACGCTCTCTGTTGGG	5833
Db	5765	CTTGGGGGATGGGTGGCTGCCAACTGCGGCTCCAGCGGCTCGGTACGCTCTCTGTTGGG	5824
Qy	5834	CGCCGGCATCGCCGAGCGGCTGTGTGGCAGCATAGGCTTTGGGAAGTGTCTGTGGACAT	5893
Db	5825	CGCCGGCATCGCCGAGCGGCTGTGTGGCAGCATAGGCTTTGGGAAGTGTCTGTGGACAT	5884
Qy	5894	CTTGGCGGCTATGGGCGAGGGTAGCGGGCGCATCTGTGTCCTTTAAGTTCATAGACGG	5953
Db	5885	TCTGGCGGTTATGGACGAGGTGGCGGCGGCTCGTGGCCCTTTAAGTTCATAGACGG	5944
Qy	5954	CGAGTGCCTCTCACCGGAGGACTGTGTCAACTTACTTCTGTGCGCATCTCTCTCTGTGTC	6013
Db	5945	CGAGTGCCTCTCACCGGAGGACTGTGTCAACTTACTTCTGTGCGCATCTCTCTCTGTGTC	6004
Qy	6014	CCTGTCTCTCGGGTCTGTGTGCGCAGCAATACTGCTCGGACGCTGGGCGCGAGAGGG	6073
Db	6005	CCTGTCTCTCGGGTCTGTGTGCGCAGCAATACTGCTCGGACGCTGGGCGCGAGAGGG	6064
Qy	6074	GGCTGTGCAGTGGATGAACCGGCTGATAGCGCTTCGCTTCGCGGGGTAAACCATGCTCCCC	6133

6065	DB	GGCTGTCCAGTGGATGAACCGGCTGTAGAGCTTGCCTCCGGGGTAATCATGTTCCCC	6124
6134	QY	TACGCACTATGTGCTTGAGAGCGACGCTGCAGACACGTTGACCTCAGATCCCTCTTAGCCT	6193
6125	DB	CACGCACTATGTGCTGTGAGCGACGCGCGAGCGCGTGTACTTCAGATCCCTCCAGCCT	6184
6194	QY	TACCATCACTCAACTGTCTGAAGGGCTCCACAGATGGATTAATGAGACATGCTCTACGCC	6253
6185	DB	TACCATCACTCAGCTGCTGAAGAGGCTCCACCAAGTGGATTAATGAAGACTGCTCCACACC	6244
6254	QY	ATGCTCCGGCTCGTGCCTAGAGGATGTTTCGGATTCGATATGCAACGCTGTGACTGACTT	6313
6245	DB	GTGTTCCGGCTCGTGGCTTAAGGGAATGTTTGGGACTCGATATGCAACGCTGTGACTGACTT	6304
6314	QY	CAAGACCTGGCTCCAGTCCCAAACTCCTTCGCCCGCGTACCGGGAGTCCCTTTCCTGTGATG	6373
6305	DB	CAAGACCTGGCTCCAGTCCCAAGCTCTTCGCCCGCAGCTACTCTGGAGTCCCTTTTTCGTG	6364
6374	QY	CCAAACGGGGTACAAGGAGCTGTGCGCGGGGAGCGGCATCATGCAAAACCACTGCCCATG	6433
6365	DB	CCAAACGGGGTACAAGGAGCTGTGCGCGGGAGACGGCATCATGCAAAACCACTGCCCATG	6424
6434	QY	CGGAGCACAGATCGCCGGACATGTCAAAAAGGTTCCATGAGGATCGTAGGGCTAGAAC	6493
6425	DB	TGGAGCACAGATCAACCGGACATGTCAAAAAAGGTTCCATGAGGATCGTCGGGCTTAAGAC	6484
6494	QY	CTGCAGCAACACGCTGGCACGGAAGTTCCCATCAACGCAATACACACGCGGACCTTCGAC	6553
6485	DB	CTGCAGCAACACGCTGGCATGGAACATTCCTCCATCAACGCAATACACACGCGGACCTTCGAC	6544
6554	QY	ACCTCTCCCGCGCGCCAACTATTCCAGGGCGGCTATGCGGGTGGCTCTGAGGAGTACGT	6613
6545	DB	ACCTCTCTCCAGCGCCAACTATTCTAGGCGCTGTGCGGGTGGCGCTGAGGAGTACGT	6604
6614	QY	GGAGGTTACCGGTGTGGGGATTTCGACTAGTGAACGGGCATGACACTGACAAACGTAAA	6673
6605	DB	GGAGGTTACCGCGGTGGGGATTTCGACTAGTGAACGGGCATGACACTGACAAACGTAAA	6664
6674	QY	GTGCCCATGCCAGTTCCGGCCCCCGAAATCTTTCACGAGGTGGATGGAGTGGGTGCA	6733
6665	DB	GTGCCCATGCCAGTTCCGGCTCTCGAATCTTCTTCGGAGGTGGAGGATGGGTGGGTGCA	6724
6734	QY	CAGGTACGCTCCGGCGTGCAAACTCTCTTCACGGGAGGACGTCAACGCTCCAGGTCGGGCT	6793
6725	DB	CAGGTACGCTCCGGCGTGCAAGGCTCTCTTACGGGAGGAGTTACATTCAGGTCGGGCT	6784
6794	QY	CAACCAATACCTGTGTGGGTGGAGTCCCATGCGAGCCCGAACCGGACCTACAGTGCT	6853
6785	DB	CAACCAATACCTGTGTGGGTGGAGTCCCATGCGAGCCCGAACCGGACCTACAGTGCT	6844
6854	QY	TACTTCCATGCTCACCGATCCCTCCACATTTACGACGACGCGCTAAGCGTAGGCTGGC	6913
6845	DB	CACCTTCATGCTCACCGACCCCTCCACATCAACGACGAAACGGCTAAGGCTAGGTTGGC	6904
6914	QY	TAGAGGCTCTCCGCCCTTTTAGCCAGCTCATCAGCTAGCCAGTTGCTCGCGCTTCTTT	6973
6905	DB	CAGGGGCTCTCCGCCCTCTCTTGCCAGCTCTTCAGCTAGCCAGTTGCTCGCGCTTCTTT	6964
6974	QY	GAAGGACATGCACTACCAACATGACTCCCGGAGCGCTGACTCATCGAGGCGCAACCT	7033
6965	DB	GAAGGACATGCACTACCAACATGCTCTCCCGAGCGCTGACTCATCGAGGCGCAACCT	7024
7034	QY	CTTGTGGCGGACGAGATGGCGGGAACATCATCTCCGCTGGATTCAGAGATPAGGTAGT	7093
7025	DB	CCTGTGGCGGACGAGATGGCGGGAACATCATCCCGCTGGATTCGAGGGAACAGGTGGT	7084
7094	QY	AATCTTGGACTCTTTTCGAAACCGCTTTCACGGGAGGGGATGAGAGGAGATATCCCGTCG	7153
7085	DB	AGTCTTGGACTCTTTTCGACCCGCTTCGAGCGAGAGGATTCAGAGGAGATATCCCGTTC	7144
7154	QY	GGCGGAGATCCTGCGAAATCCAGGAAGTTCCCTCTCAGCGCTCCCATATGCGCACGCC	7213

Qy	9374 AACGGGGAGCTAACCACTCCAGGCGCTAAGGCATTTCCGTGTTTTTTTTT	9425
Db	9365 AACGGGGAGATAACACTCCAGGCCAATAGGCCATCCCGCTTTTTTTTTTTT	9416

```

RESULT 3
US-08-384-616-1
; Sequence 1, Application US/08384616
; Patent No. 5847101
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeod &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/384,616
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Stevens-Smith, Theresa M.
; REGISTRATION NUMBER: 36,281
; REFERENCE/DOCKET NUMBER: 900703B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; TELEX: 440142
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA from genomic RNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 333..9362
; PS-08-384-616-1

```

QY	14	TGGGGCGACACTCCACCATTGAATCACTCCCTGTGAGGAACACTACTGTCTTCACGAGAA	73
Db	5	TGGGGCGCACACTCCACCATTAGATCACTCCCTGTGAGGAACACTACTGTCTTCACGAGAA	64
QY	74	AGCGTTAGCCATGGCGTTAGTATGATGATGCTGCTGACAGCTTCAGGACCCCCCTTCCCGGG	133
Db	65	AGCGTTAGCCATGGCGGTTAGTATGATGATGCTGCTGACAGCTTCAGGACCCCCCTTCCCGGG	124
QY	134	AGAGCCATAGTGCTTGCGGAAACCGGTGAGTACACCGGAATTCGCCAGGACGACCGGGTCC	193
Db	125	AGAGCCATAGTGCTTGCGGAAACCGGTGAGTACACCGGAATTCGCCAGGACGACCGGGTCC	184
QY	194	TTTCTTTGGATCAACCCCGCTCAATGCCTGGAGAATTGGGCGTGCCCCGCGAGACTGCTAG	253
Db	185	TTTCTTTGGATCAACCCCGCTCAATGCCTGGAGAATTGGGCGTGCCCCGCGAGACTGCTAG	244
QY	254	CCGAGTAGTGTTGGGTTCGGAAGGCTTTGTGTACTGCTGTATAGGTGCTTTGCCAGTG	313
Db	245	CCGAGTAGTGTTGGGTTCGGAAGGCTTTGTGTACTGCTGTATAGGTGCTTTGCCAGTG	304
QY	314	CCCCGGAGGTCGCTAGACCGGTGCACATGACGAACTCTAAACCTCAAAGAAAAAC	373
Db	305	CCCCGGAGGTCGCTAGACCGGTGCACATGACGAACTCTAAACCTCAAAGAAAAAC	364
QY	374	CAAACTAACACAAACCGCGGCCACAGGACGTCAAGTTCCCGGGCGGTGGTCAGATCGT	433
Db	365	CAAACTAACACAAACCGCGGCCACAGGACGTCAAGTTCCCGGGCGGTGGTCAGATCGT	424
QY	434	TGCTGAGATTACCTTTTCCCGCGAGGGCCCCACAGTTGGGTGTCGGCGCATAGGAA	493
Db	425	TGCTGAGATTACCTTTTCCCGCGAGGGCCCCACAGTTGGGTGTCGGCGCCCCAGGAA	484
QY	494	GGCTTCCGAGCGGTCGCAACCTCGTGGAAAGCGACAACTATCCCAAAGGCTCGCCGACC	553
Db	485	GACTTCCGAGCGGTCGCAACCTCGTGGAAAGCGACAACTATCCCAAAGGCTCGCCGACC	544
QY	554	CGAGGCGAGGGGCGTGGGCTCAGCCCGGGTACCTTTGGCCCCCTCATGGCAATGAGGACCT	613
Db	545	CGAGGCGAGGACCTGGGCTCAGCCCGGGTACCTTTGGCCCCCTCATGGCAATGAGGACCT	604
QY	614	GGGTTGGGACAGATGACTCTGTCAACCGCGGCTCCCGGCTTAGTTGGGGCCCCACGGA	673
Db	605	AGGTTGGGACAGATGACTCTGTCAACCGCGGCTCCCGGCTTAGTTGGGGCCCCACGGA	664
QY	674	CCCCCGGCTAGGTCGCTAACTTGGGTAAAGTTCATCGATACCCCTTACATCGCGCTTCGC	733
Db	665	CCCCCGGCTAGGTCGCTAACTTGGGTAAAGTTCATCGATACCCCTTACATCGCGCTTCGC	724
QY	734	CGATCTCATGGGTTACATTCGCTCTGTGGCGCCCCCTTAGGGGGCGCTGCCAGGACCTT	793
Db	725	CGATCTCATGGGTTACATTCGCTCTGTGGCGCCCCCTTAGGGGGCGCTGCCAGGACCTT	784
QY	794	GGCACACGGTTCGCGGTTCTGGAGACGGCTGAACTATGCAACAGGGAACCTTCCCGCG	853
Db	785	GGCACATGGTTCGCGGTTCTGGAGACGGCTGAACTATGCAACAGGGAACCTTCCCGCG	844
QY	854	TTGCTCTTTCTCTATCTTCTCTTGGCTCTGCTGTCTTGTGAACATCCCAAGTTCGCG	913
Db	845	TTGCTCTTTCTCTATCTTCTCTTGGCTCTGCTGTCTTGTGAACATCCCAAGTTCGCG	904
QY	914	TTATGAAGTGCACAACGCTGCCGGATATACCATGTTCAGAAACGACTGTCTCAACTCAAG	973
Db	905	TTATGAAGTGCACAACGCTGCCGGATATATCATGTTCAGAAACGACTGTCTCAACGCAAG	964
QY	974	CATTGTGTATGAGGACAGCGGACGTGATCATATGCANATCTCCGGGTGGTCCCTGTGTTCA	1033
Db	965	CATTGTGTATGAGGACAGCGGACTTGTATCATATGCATCTCTGCGGTGGTCCCTGTGCTG	1024
QY	1034	GGAGGTTACAGCTCCGTTGTCTGGTAGGCTCACTCCACGCTCGCGGACAGATATGC	1093
Db	1025	GGAAGGCAACTCCTCCGCTGTCTGGTAGGCTCACTCCACGCTCGAGACAGGAACGT	1084
QY	1094	CAGCGTCCCACATTACGACAAATACGACGCAACGTCGACTTGTCTGTGGACGGCTGCTT	1153

1085 Db CACCATCCCCACACGACGATACGACGCCAGCTGATCTGCTCGTGGGCGGCTGCTTT 1144
1154 QY CTGCTCGGTATGTAAGCTGGGGGATCTCTGGGATCTATTTCTGCTCTCCACAGCTGTT 1213
1145 Db CTGTTCCGCTATGTACCTGGGGGACCTCTCTGGGATCTGTTTCTGCTCTCAGCTGTT 1204
1214 QY CACCTTCTCGCTCCGCGGATGAGACAGTGCAGGACTGCAACTGCTCAATCTATCCCGG 1273
1205 Db CACCTTCTCGCTCCGCGGATGATGATATACAGGACTGTAACTGCTCAATTTATCCCGG 1264
1274 QY CCAATGATCAGGTCACCGCATGGCTTGGGATATGATGATGAATGAATGCTCAACACAGC 1333
1265 Db CCAATGATCAGGTCACCGCATGGCTTGGGATATGATGATGAATGAATGCTCAACACAGC 1324
1334 QY CCTAGTGGTGGAGTCTCGGATCCGATCCACAGGCTGCTGGACATGTCGGGGG 1393
1325 Db CCTAGTGGTGGAGTCTCGGATCCGATCCACAGGCTGCTGGACATGTCGGGGG 1384
1394 QY CCATGGGGAGTCTCGGCGGCTTGCCTACTATTCATGATGGTGAACCTGGGCTAAGGT 1453
1385 Db CCATGGGGAGTCTCGGCGGCTTGCCTACTATTCATGATGGGGAACCTGGGCTAAGGT 1444
1454 QY TCTGATTTGGCGTACTCTTTCGGCGCTTGACGGGAGACCCACAGCGGGAGGGT 1513
1445 Db TCTGATTTGGCGTACTCTTTCGGCGCTTGACGGGAGTACCCACGCTGACAGGGGGG 1504
1514 QY GGCCTGGCCACACACCTCCGGTTCAGCTCCCTTTCTCATCTGGGCGCTCAGAAAT 1573
1505 Db GCAAGCCAAACACACAGGCTGCTGTCATGTTCCGAGTGGGCGCTCAGAAAT 1564
1574 QY CCGAGTCTGTAATACCAACGGGAGCTGGGACATCAACAGGAGCTGCCATAATTCGAATGA 1633
1565 Db CCGAGTCTGTAATACCAACGGGAGTGGGACATCAACAGGAGCTGCCCTGAATGCAATGA 1624
1634 QY CTCCTCCAACTGGTCTTTCGGCGCTGTTTACGGACACAAAGTTCGAATGCTGCCG 1693
1625 Db CTCCTCCAACTGGTCTTTCGGCGCTGTTTACACACATAGTTTCAACTCGTCCGG 1684
1694 QY GTGCCCGGAGGATGAGCCAGCTGCCCGCCCATGACTGCTTCCGCCAGGGTGGGGCC 1753
1685 Db GTGCCAGAGGATGAGCCAGTGCACGACCATGACAGTTGACAGGGATGGGGTCC 1744
1754 QY CATCACTATPATAAGCTTAAGCTCGATGAGGCTTATGCTGGCAATACGCGCC 1813
1745 Db CATTACTATGCTGAGCTAGCAGATCAGACACAGAGGCCATATTGCTGGCACTACCCACC 1804
1814 QY TCAGCCGTGTGTGCTATCCCGCTCGAGGTGTGTGCTCGAGTGTATTTGTTTCACCCC 1873
1805 Db TCACAAATGTACCATGTACTCTGCTGGAGGTGTGCGGCCCGAGTGTACTGCTTCACCCC 1864
1874 QY AAGCCCTGTTGTGGTGGGACCCGATCGTTCCGGTGTCCCTACGATATAGCTGGGGGA 1933
1865 Db AAGCCCTGTTGTGGTGGGACCCGATCGTTCCGGTGTCCCTACGATATAGTGGGGGA 1924
1934 QY GAATGACAGAGTGTATGCTCTCAACACACGCTCGGCTCGGACAGGCAACTGTTGCG 1993
1925 Db GAACGAGAGTGTATGCTGCTGTCTCAACACACGCTCGGCTCGGACAGGCAACTGTTGCG 1984
1994 QY CTGTACATGATCAATAGTACTGGTTCATTAAGAGCTGGAGGTCCTCCCGCTTAACAT 2053
1985 Db CTGCACATGATGAATAGCACCGGGTTCCACACACATGTTGGGGGGCCCCCGTGTAACT 2044
2054 QY CGGGGGGTGGTAAACCGCACTTGATCTGCCCAACGACTGCTTCGGGAAGCACCCCGA 2113
2045 Db CGGGGGGTGGTAAACCGCACTTGATCTGCCCAACGACTGCTTCGGGAAGCACCCCGA 2104
2114 QY GGTACTTACAAAATGTGGCTTGGGGCTTGGGCTGACACTAGTGTGCTAGACTA 2173
2105 Db GGTACTTACAAAATGTGGTTCGGGGCTTGGGCTGACACTAGTGTGCTAGACTA 2164
2174 QY CCCATACAGGCTTTGGCACTACCCCTGCACTCTCAATTTTCCATCTTTAAGGTTAGGT 2233

2165 Db TCCATAAGGCTCTGGCATTAACCCCTGCACTGTAACTTTAACAATCTCAAGTTAGGAT 2224
2234 QY STATGTGGGGGGGTGGACACAGGCTCAATGCGGATGCAATTTGGACTTCGAGGAGCG 2293
2225 Db STATGTGGGGGGGTGGACACAGGCTCAATGCGGATGCAATTTGGACCCGAGGAGCG 2284
2294 QY CTGTAACTTTGGAGGACAGGATAGGTCAAACTCAGCCCGCTGCTGCTCTPACAAACAGA 2353
2285 Db TTGTGACTTTGGAGGACAGGATAGGACCGGAGCTCAGCCCGCTGCTGCTCTPACAAACAGA 2344
2354 QY GTGGCAGATPACTGCCCCTGTGCTTTTCAACCCCTACCGGCTTTATCCACTGTTGATCCA 2413
2345 Db GTGGCAGGTAATGCCCCTGTGCTTTTCAACCCCTACCGGCTTTTCACTGGCTGATTTCA 2404
2414 QY TCTCCATCAGACATCGTGGAGTGCATACCTGTACGCTAGGCTAGGCTAGGCTTTCTCTC 2473
2405 Db CTTCCATCAGACATCGTGGAGTGCATACCTGTACGCTAGGCTAGGCTTTCTCTC 2464
2474 QY CTTTGCNAATCAATGGGAGTACATCTGTTGCTTTTCTCTCTCTGGCAGACGCGCGCT 2533
2465 Db CTTTGCNAATCAATGGGAGTGTCTGTTGCTTTTCTCTCTCTAGCGSACGCACTGT 2524
2534 QY GTGTGCTCTGTTGATGATGCTGCTGATAGCCAGGCTAGGCGCTTACAGAACTT 2593
2525 Db GTGTGCTCTGTTGATGATGCTGCTGATAGCCAGGCTAGGCGCTTACAGAACTT 2584
2594 QY GGTGTCTCTCAATGGGCGTCCGTGGCGGAGCGATGATTTCTCTCTCTCTCTGTTGTT 2653
2585 Db GGTGTCTCTCAATGGGCGTCTGTGGCGGCGGCAATGGCATCTCTCTCTCTCTGTTGTT 2644
2654 QY CTTTTCGCGCGCTGATCAATTAAGGCGAGGTGGCTCCTGGGGCGGCTGATGCTTTTGA 2713
2645 Db CTTTTCGCGCGCTGATCAATTAAGGCGAGGTGGCTCCTGGGGCGGCAATGCTCTTTA 2704
2714 QY TGGCGTATGCGCGCTGCTCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2773
2705 Db TGGCGTATGCGCGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2764
2774 QY CCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2833
2765 Db CCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2824
2834 QY GTCACTACTACTCAAAAGTGTCTCTCACTAGGCTCATATGGTGTACAAATCTTTATCAC 2893
2825 Db GTCACTACTACTCAAAAGTGTCTCTCACTAGGCTCATATGGTGTACAAATCTTTATCAC 2884
2894 QY CAGAGCGGAGGCGACATGCAAGTGTGGTCCCCCCCCCTCAACGTTCCGGGAGCGCGGA 2953
2885 Db CAGAGCGGAGGCGACTTACATGTGTGAATCCCCCCCCCTCAACGTTCCGGGAGCGCGGA 2944
2954 QY TGCATCATCTCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3013
2945 Db TGCATCATCTCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3004
3014 QY CTTGCTGCGCACTCTGCGGCCCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3073
3005 Db TCTAATTTGCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3064
3074 QY CTTGCTGCGCTCAAGGCTCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3133
3065 Db CTTGCTGCGCTCAAGGCTCATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3124
3134 QY TCATTAATGCTCAAAATGCTTCTCATGAAGTGGCGGCTGACAGGATGCTAGCTTTATAA 3193
3125 Db TCATTAATGCTCAAAATGCTTCTCATGAAGTGGCGGCTGACAGGACGCTACATTTACA 3184
3194 QY CCAATCTTACCCCACTGCGGGAATGCGGCCCAACGCGGCTTACGAGACCTTGGGCTGGCGGT 3253
3185 Db CCAATCTTACCCCACTGCGGGAATGCGGCCCAACGCGGCTTACGAGACCTTGGGCTGGCGGT 3244
3254 QY AGAGCCCTGCTCTCTCGCCATGAGGACGAGCTGCTCATCTCTGCGGAGGACGACCGC 3313
3245 Db AGAGCCCTGCTCTCTCGCCATGAGGACGAGCTGCTCATCTCTGCGGAGGACGACCGC 3304

QY	3314	TGCGTGTGGGGACATCATCTTTGGGTCTACCGGTCTCGCCCGAAGGGGGGAAGGAGATATT	3373
Db	3305	GGCGTGTGGGACATCATTTGGGTCTCGCGGTCTCGCCCGAAGGGGAAGGAGATACT	3364
QY	3374	TTTGGGACCGGCTGATAGTCTCGAAGGGCAAGGGTGGCGACTCTCTTCGCGCCCATTCACGGC	3433
Db	3365	CTTGGGCGCGGCGCATAGTCTTGAAGGGCGGGGTTCGCACTCTCTCGCGCCCATTCACGGC	3424
QY	3434	CTACTCCCAACAAACGCGGGGCGTACTTGGTTGTCATCATCACTAGCCTCACAGGCCGGA	3493
Db	3425	CTACTCCCAACAGAGCGGGGCGTACTTGGTTGTCATCATCACTAGCCTCACAGGCCGGA	3484
QY	3494	CAAGAACAGGTGGAAGGGGAGGTTCAAGTGGTTTCTACCGCAACACAATCTTTCTGTGGC	3553
Db	3485	CAAGAACAGGTGAGGGAGGTTCAAGTGGTTTCCACCGCAACACAATCTTTCTGTGGC	3544
QY	3554	GACCTGCATCAACCGCGGTGTGCGACTGTCTACCATGGCGCTGGCTCGAAGACCCCTAGC	3613
Db	3545	GACCTGTGTCAACGGCGGTGTGCGACGTTTACCATGGTGTGGCTCAAGACCTTTAGC	3604
QY	3614	CGGTCCAAAAGGTCNAAATCAACCAATGTACACCAATGTAGACTGTGACCTCGCTCGGCTG	3673
Db	3605	CGGCCAAAAGGGCCATCACCCAGATGTACTAAATGTGGAACAGGACCTCGTGGCTG	3664
QY	3674	GCAGGCGCCCCCGGGGGCGCTCCATGACACCATGAGAGTGTGGAGCTTCGAGCACTTTA	3733
Db	3665	GCCCAAGCCCCCGGGCGGCTTCTTTGACCACTGACCGTGTGGAGCTTCAGACCTTTA	3724
QY	3734	CTTGGTCAAGAGACATGCTGATGTCAITTCGGGTGCGCGCGAGGGGCACAGCAGGGAAG	3793
Db	3725	CTTGGTCAAGAGACATGCTGAGGTCAITTCGGGTGCGCGCGAGGGGCACAGTAGGGGAG	3784
QY	3794	TCTACTCTCCCCAGGCGCGTCTCTACCTGAAGGCTCTCTCGGGTGGTCCATTCGCTTTG	3853
Db	3785	CTGTCTCTCCCCAGGCGTGTCTCTTACTTGAAGGCTCTCTCGGGTGGTCCATTCGCTTTG	3844
QY	3854	CCCTTCGGGACACGTCTGTGGGCGTCTTCGGGCTGTGTGTGACCCCGGGGGTTCGCA	3913
Db	3845	CCCTTCGGGACACGTCTGTGGGCGTCTTCGGGCGTGTGTGTGACCCCGGGGGTTCGCA	3904
QY	3914	GGCGGTGGACTTCATACCCGCTTGAGTCTATGGAATACCATGGGTCTCGGGTCTTCAC	3973
Db	3905	GGCGGTGGACTTTGTGCCCCGTAGAGTCCATGGAACTACTATCGGTCTCGGGTCTTCAC	3964
QY	3974	AGACAACTCAACCCCCCGGCTGTACCGCAGACATTCGAAAGTGGCACATCTGCACGCTCC	4033
Db	3965	GGACAACTCATCCCCCGCGCGTACCGCACTCATTTCAAGTGGCCACCTACACGCTCC	4024
QY	4034	TACTGGAGCGGCAAGAGCACTCAAGTTCGCGTGGTATGCAAGCCCAAGGTACAAAGT	4093
Db	4025	CACCTGGAGCGGCAAGAGTACTAAGTTCGCGGTGCATATGCAAGCCCAAGGTACAAAGT	4084
QY	4094	GCTCGTCTCAACCCGCTCGCGCCACTTAGGGTTTGGGCGGTATATGTGCAAGGC	4153
Db	4085	GCTCGTCTCAATCCGTCGCTTGGCGTACTTAGGGTTTGGGCGGTATATGTGCTAAGGC	4144
QY	4154	ACAGGTATCGACCCCAACTCAGAACTGGGGTAAGGACCATTAACACGGGCGGCTCCAT	4213
Db	4145	ACAGGTATTGACCCCAACTCAGAACTGGGGTAAGGACCATTAACACAGGCGCCCGGT	4204
QY	4214	TAGGTACTCAACTATGGCAAGTTCCCTTTCGCGAGGTGGGTGTCTTGGGGCGGCTTATGA	4273
Db	4205	CACATACTCTACCTATGGCAAGTTCTTTCGCGATGGTGTGCTCTGGGGCGGCTTATGA	4264
QY	4274	CATCATATATGTGATGAGTGCCACTCAACTGATCGACTACCATCTTGGGCACTCGGCAC	4333
Db	4265	CATCATATATGTGATGAGTGCCACTCAACTGATCGACTACCATCTTGGGCACTCGGCAC	4324
QY	4334	AGTCTCTGCAACCAAGCGGAGACGGTGGAGCGGCGTCTGTGCTCGGCACCGCTACAC	4393
Db	4325	AGTCTCTGCAACCAAGCGGAGACGGCTGGAGCGGCGGCTTGTGCTGCTGCCACCGCTACGC	4384

Qy	4394	TCGGGATCGGTACCGTGGCCACACCCCAATATATGAGAAATAGGCTGTGCCAACAATGG	4453
Db	4385	TCGGGATCGGTACCGTGGCCACACCCAAACATCGAGAGTGGCCCTGTCTAATACTCGG	4444
Qy	4454	AGAGATCCCTCTTATGGCAAGCCATCCCCATTGAGGCCATCAAGGGGGGAGGCGATCT	4513
Db	4445	AGAGATCCCTTCTATGCGAAAGCCATCCCCATTGAGCCCATCAGGGGGGAAAGGCATCT	4504
Qy	4514	CATTTCCTGCCATTCCAAGAGAAATGTGACGAGCTCGCCGCAAGGTGACAGGCCCTCGG	4573
Db	4505	CATTTCCTGTCATTCCAAAGAAAGTGGCAGCAGCTCGCCGCAAGCTGTCAAGCCCTCGG	4564
Qy	4574	ACTCAACGCTGTAGCATATTACCGGGGCCCTTGATGTGTCTCGGTCAATACCGCCTATCGGAGA	4633
Db	4565	AATCAACGCTGTGGCGTATTACCGGGGGCTCGATGTGTCTCGGTCAATCAACCACTATCGGAGA	4624
Qy	4634	CGTCGTTGCTGGTGAACACAGCCTCTAATGACGGGTTTCACCGGCGATTTTGACATCAGT	4693
Db	4625	CGTCGTTGCTGGTGAACACAGCCTCTGATGACGGGCTATACGGGGCGACTTTTGACATCAGT	4684
Qy	4694	GATCGACTGCAATATACATGTGTCAACCAGACAGTCGACTTCAGCTTGGATCCCACTTTCAC	4753
Db	4685	GATCGACTGTAAACATGTGTCTACCCAGACAGTCGACTTCAGCTTGGATCCCACTTTCAC	4744
Qy	4754	CATTGAGACGACCGTGCCTCCCAAGACGCGGTGTGCGCTTCGCAACGCGGAGGTGAAC	4813
Db	4745	CATTGAGACGACCGTGCCTCAAGACGCGGTGTGCGCTTCGCAACGCGGCGGTGAGAC	4804
Qy	4814	TGCAGGGGTAGAGTGGCATCTACAGGTTTGTGACTCCAGGAGAACGCGCCCTCGGGCAT	4873
Db	4805	TGGCAGGGGTAGGAGAGGCATCTACAGGTTTGTGACTCCGGGAGAACGCGCCCTCGGGCAT	4864
Qy	4874	GTTTCGATTCTTCGTCCTGTGTGAGTGTCTAATGACGCGGCTGTGCTTGGTATCAGCTCAC	4933
Db	4865	GTTTCGATTCTTCGTCCTGTGTGAGTGTCTAATGACGCGGCTGTGCTTGGTATCAGCTCAC	4924
Qy	4934	GCCCGCTGACCTCGGTAGTTGGGGCTTACCTAAATACACACCGGTTGCCGCTG	4993
Db	4925	CCGGCCGAGACCTCGGTTAGTTGGGGCTTACCTGAAACACACCGGTTGCCGCTG	4984
Qy	4994	CCAGGACCATCTGGAGTCTGGGAGAGCGTCTTTCACAGGCTCACCCCAATATATGCCCA	5053
Db	4985	CCAGGACCATCTGGAGTCTGGGAGAGTGTCTTTCACAGGCTCACCCCAATATATATGCCCA	5044
Qy	5054	CTTCTCTGCCAGACTAAACACGACGAGACAACTTTCTTACCTGGTGGCAATCAAGC	5113
Db	5045	CTTCTCTGCCAGACTAAACACGACGAGAGCACTTTCCCTTACTCTGGTAGCATCAAGC	5104
Qy	5114	TACAGTGTGCGCAGGCTCAAGCTCCACTCCATCGTGGGACCAATGTGGAAGTGTCT	5173
Db	5105	CACGTTGTGCGCAGGCTCAGGCCCACTCCATCATGGGATCAAAATGTGGAAGTGTCT	5164
Qy	5174	CATAAGCTGAARACTTACACTGACCGGCCAACACCCCTGCTGTATAGCTTAGGAGCCGT	5233
Db	5165	CATAAGCTGAARACTTACACTGACCGGCCAACACCCCTTGTGTATAGCTTAGGAGCCGT	5224
Qy	5234	CCAAATAGAGTGCATCTCTACACACCCCAATACTAAATACATCATGGCATGATGTCGGC	5293
Db	5225	CCAGATAGGTCACTCTACACCCCAATACTAAATACATCATGGCATGATGTCGGC	5284
Qy	5294	TGACCTGGAGTTCGTCACTAGCACTTGGGTGTGGTAGCGGAGTCCCTTGACGCTTTGCG	5353
Db	5285	TGACCTGGAGTTCGTCACTAGCACTTGGGTGTGGTAGCGGAGTCCCTTGACGCTTTGCG	5344
Qy	5354	CGCATACTGCTGACGACAGGCGAGTGTGTCTATTGTGGGAGGATCATCTTGTCCGGAA	5413
Db	5345	CGCGTATGCTCTACAAACAGGCGAGTGTGTCTATTGTGGGTAGGAATTATCTTGTCCGGAG	5404
Qy	5414	GCCAGCTGTCTTCCCGACAGGGAAGTCTCTACAGGAGTTCGATGAGATCGAGAGTGT	5473
Db	5405	GCCGGCATTTGTTCCCGACAGGAGTCTCTCTACAGGAGTTCGATGAAATGAAGAGTGT	5464
Qy	5474	TGCTCTCAACACTTCTCTTATCATCGACAGGGAATGACAGCTCGCGAGCAATTCAGCAAAA	5533

Db	5465	CGCTCGACCTCCCTTACATCGAGCAGGGAATCAGTCCGAGCAATCAAGCAGAA	5524
Qy	5534	GGCCTCGGGTGTGTGAACCGGCCACCAAGCAACGAGAGCTGCTGCTCCCGTGTGG	5593
Db	5525	AGCGCTCGGGTTACTGTCAAAACAGCCACCAACAGCAGGAGCTGCTGCTCCCGTGTGG	5584
Qy	5594	GTCCAAGTGGCAGCGCTTGAGACCTTCTGGGCGAAGCAATGTGGAATTTCTACGCG	5653
Db	5585	GTCCAGTGGGAGCCCTTGAGACATTTCTGGGCGAAGCAATGTGGAATTTCTACGCG	5644
Qy	5654	AATACAGTACCTAGCAGGCTTATCCACTCTGCCTCGAAACCCGCGATAGCATCATGAT	5713
Db	5645	GATACAGTACTTTAGCAGGCTTATCCACTCTGCCTGGAAACCCGCGATAGCATCATGAT	5704
Qy	5714	GGCATTTACAGCTTCTATCACTAGCCGCTCACCAACCAAAACACCTCTCTGTTTAAAT	5773
Db	5705	GGCAATCACAGCTCTATCACAGCCGCTCACCAACCAAGTACCTCTCTGTTTAAAT	5764
Qy	5774	CTTGGGGGATGGTGGCTGCCAACTCGTCTCTCCAGCGCTCGCTCAGCTTCGTGGG	5833
Db	5765	CTTGGGGGGTGGTGGTCCCAACTCGCCCCCCAGCGCGCTTCGCGTTCGTGGG	5824
Qy	5834	CGCCGGATCGCCGAGCGGCTGTTCGACAGATAGGCTTTGGGAAGTGTCTGTGACAT	5893
Db	5825	CGCCGGATCGCCGCTGGTGGCAGATAGGCTTTGGGAAGTGTCTGTGACAT	5884
Qy	5894	CTTGGCGGCTATGGGCGAGGGTAGCGGCGACCTGCGCTTAAAGTCTATGACGG	5953
Db	5885	TCYGGCGGTTATGGACAGAGTGGCGCGCGCTCGTGCCCTTAAAGTCTATGACGG	5944
Qy	5954	CGAGTCCCTTCCACCGAGCACTGTGCAACTTACTTCCCTGCCATCCTCTCTCTGCTGC	6013
Db	5945	CGAGATGCCCTCCACCGAGCACTGTGCAACTTACTTCCCTGCCATCCTCTCTCTGCGC	6004
Qy	6014	CTTGGTGGTGGGCTGTGTCGACAGCAATCTGCTGCGGCACTGTGGGCGCGAGGG	6073
Db	6005	CTTGGTGGTGGGCTGTGTCGACAGCAATCTGCTGCGGCACTGTGGGCGCGAGGG	6064
Qy	6074	GGCTGTGCAGTGAATGAACCGGCTGTAGAGGTTGCTTCCGCGGGTAAACCACTCTCCC	6133
Db	6065	GGCTGTGCAGTGAATGAACCGGCTGTAGAGGTTGCTTCCGCGGGTAAACCACTCTCCC	6124
Qy	6134	TAGCATATGTGCCTGAGAGCGAGCTGACGACGCTGACCTCAGATCCTCTCTAGCCT	6193
Db	6125	CAGCACTATGTGCTGAGAGCGACGCGCAGAGCGGTGTACTCAGATCCTCTCTAGCCT	6184
Qy	6194	TACCATCACTCAACTGCTGAAGCGGCTCCACAGTGAATTAATGAGACGTCTCTAGCC	6253
Db	6185	TACCATCACTCAGCTCTGAAGAGGCTCCACAGTGAATTAATGAGACGTCTCTAGCC	6244
Qy	6254	ATGCTCCGGCTGTGGCTTAAGGATGTTTGGGATTTGGATATGACCGGTGTGACTGACTT	6313
Db	6245	GTGTCCGGCTGTGGCTTAAGGATGTTTGGGATTTGGGATTTGACCGGTGTGACTGACTT	6304
Qy	6314	CAAGACTGCTCCAGTCCAAAATCTCTGCGCGGTTTACCGGAGTCCCTTCTCTGCTAG	6373
Db	6305	CAAGACTGCTCCAGTCCAAAATCTCTGCGCGGTTTACCGGAGTCCCTTCTCTGCTAG	6364
Qy	6374	CCAAACCGGGTAAAGGAGTCTTGGCGGGGAGCGGATCATGCAAAACCACTTCCCATG	6433
Db	6365	CCAAACCGGGTAAAGGAGTCTTGGCGGGGAGCGGATCATGCAAAACCACTTCCCATG	6424
Qy	6434	CGAGACACAGATCGCCGGAATGTCAAAAACGGTTCATCAGGATCGTAGGGCTTAGAAC	6493
Db	6425	TGGAGCACAGATCACCGGAATGTCAAAAACGGTTCATCAGGATCGTAGGGCTTAGAAC	6484
Qy	6494	CTGCAGCAACAGTGGCAGCGAAGTTCGCCATCAACGCAATACACGAGGACCTTGGCAC	6553
Db	6485	CTGCAGCAACAGTGGCAGCGAATTTCCCAATCAAGCGAATACACGAGGACCTTGGCAC	6544
Qy	6554	ACCTTCCCGCGGCCAACTATTTCAGGGCGCTATGGCGGGTGGCTGCTGAGGATACGT	6613

D5	6345	ACCCTCTCAGCGCCAAACTATTCTTAGGGCGCTGTGGCGGGTGGCCGCTGAGGAGTACGT	6604
Q5	6614	GGAGGTATCAGCGTGTGGGGGATTTCCACTACGTGCA CGGCGATACACACTGACAACTGATAA	6673
D5	6605	GGAGGTCAACGCGGTGSGGGATTTCCACTACGTGCA CGGCGATGACCACTGACAACTGATAA	6664
Q5	6674	GTGCCCATGCAAGTTCCGGGCCCGGAATTCCTTCA CGGAGGTGGAATGGAGTGCCTTGCAT	6733
D5	6665	GTGCCCATGCAAGTTCCGGGTCTCGAATTCCTTCT CGGAGGTGGA CGGAGTGCCTTGCAT	6724
Q5	6734	CAGGTACGCTCCGGCGTGC AAACTCTCTTACCGGAGGACGTCA CGTTCAGTTCAGTCTCGGCT	6793
D5	6725	CAGGTACGCTCCGCGTGCAGGCTCTCTTACCGGAGGAGTTCATTCAGGTCTCGGCT	6784
Q5	6794	CAACCAATATTGGTTCGGGTTCGAGCTCCCAATGCGAGCCCGGAACCGGACGTAACTACGTCT	6853
D5	6785	CAACCAATATTGGTTGGGTCA CAGCTTACATGCGAGCCCGGAACCGGATGACAGCTGCT	6844
Q5	6854	TACTTCATGCTCACCGATCCCTCCCACTTACAGACGAGACGCGCTAAGCGTAGGCTGGC	6913
D5	6845	CATTTCATGCTCACCGACCCCTCCCACTACAGACGAGAAACCGCTAAGCGTAGGTTGGC	6904
Q5	6914	TAGAGGCTCTCCGCCCTCTTTAGCCAGCTCATCAGCTAGCCAGTTGTTCTTCGCGCTCTTTT	6973
D5	6905	CAGGGGCTCTCCGCCCTCTTGGCCAGCTCTTCAGCTTAGCCAGTTGTTCTTCGCGCTCTCTT	6964
Q5	6974	GAAGGCACATGCACTACCAACATGCTCTCCCGGAGCTGACCTCATGAGAGCCCACT	7033
D5	6965	GAAGGCACATGCACTACCAACCATGCTCTCTCCCGAGCGTGA CTTCATGAGGCCAACCT	7024
Q5	7034	CTTTGTCGGCAGAGATGGCGGAAACATCACTCGCGTGGAGTTCAGAGAAATAGGTAGT	7093
D5	7025	CCTGTGCGCAGGAGATGGCGGGAACATCA CCGCGTGGAGTCGGAGAACAGGTGCT	7084
Q5	7094	AAATTCTGGACTTTTGGAAACCGTTTACGCGGAGGGGAGTAGAGGAGAGATATCCGTGCG	7153
D5	7085	AGTCCTCGACTCTTTTGGACCCGCTTCGAGCGGAGGAGGATGAGAGGAGAGATATCCGTGCC	7144
Q5	7154	GGCGAGATCTCGGAAATTCAGGAGTTGCCCTCAGCGTTGCCCATATGGGCACGCCCC	7213
D5	7145	GGCGAGATCTCGGGAATTCAGAGAGTTCCCGCGACGAGTCCCATCTGSGCGGCGCCC	7204
Q5	7214	GGACTCAATCTCTCACTGCTTAGAGTCTTGGAAGGACCGGAGCTAGTCCCTCCGCTGGT	7273
D5	7205	GGATTAACAACCTTCACTGTAGTCTTGGAAGGACCGGAGCTAGTCCCTCCGCTGGT	7264
Q5	7274	ACCGGATGCCATGTCACCTACGAGGCTCTCCAAATACACACTCCGAGGAGAGAG	7333
D5	7265	GCACGGGTGCCGTTGCCACTATCAAGGCCCTCTCAATACCACTCCACGHAAGAG	7324
Q5	7334	GACGGTTGCTCTGACAGAAATCCAAATGTGTTCTGTCTTGGCGGAGCTGCCACTAAGAC	7393
D5	7325	GACGGTTGCTCTAACAGAGTCTCCGCTGTTCTGTGCTTAGCGAGCTCGCTACTAAGAC	7384
Q5	7394	CTTCGGTAGCTCCGGATCGTCGCGCTTGATAGGCGACGGGACCGCGCTTCTGACCT	7453
D5	7385	CTTCGGCAGCTCCGAATCATCGGCGGTGACACAGGGACGGGACCGCGCTTCTGACCA	7444
Q5	7454	GGCTCCGACGAGGTGACAAAGATCCGACGTTGAGTCGTACTCTCTCCATGCCGCCCT	7513
D5	7445	GGCTCCGACGAGGTGACAAAGATCCGACGTTGAGTCGTACTCTCTCCATGCCGCCCT	7504
Q5	7514	TGAAGGGAGCCCGGGGACCCCGATCTCAGCGACGGGTCTTGCTTACCGTAGTAGGA	7573
D5	7505	TGAAGGGGAAACCGGGGACCCCGATCTCATGTGACGGGTCTTGCTTACCGTAGTAGGA	7564
Q5	7574	GGCTAGTAGAGATGTCTGTCTGTCTCAATGTCTTATAGTGGACAGGCGCTGATCAC	7633
D5	7565	AGCTAGTAGAGATGTCTGTCTGTCTCAATGTCTTACATGACAGGCGCTTGAATCAC	7624
Q5	7634	GCCATCGCTGCGGAGGAAGTAGCTGCCCACTAACCCGTTGAGCACTCTTTGCTGG	7693
D5	7625	GCCATCGCTGCGGAGGAAGCAAGCTGCCCACTAACCCGTTGAGCACTCTTTGCTGG	7684

QY 7694 TCACCAACATGGTCTAGCCACACACATCCCGCAGCGCAAGCCTCCGSCAGAGAGGT 7753
DB 7695 CACCAATACATGGTTTATGCCACACATCTCGAGCGCAGCCTCGGCGAGAGAGGT 7744
QY 7754 CACCTTTGACAGATGCAAGTCTCGATGATCATTTACCGGGAGCTACTCAAGAGAGATGAA 7813
DB 7745 CACCTTTGACAGATGCAAGTCTCGATGATCATTTACCGGGAGCTACTCAAGAGAGATGAA 7804
QY 7814 GCGGAGGCGTCCACAGATTAAGCTTAAGCTTCTATCTATAGAGAGGCTCGAAGTGC 7873
DB 7805 GCGGAGGCGTCCACAGATTAAGCTTCTATCTATAGAGAGGCTCGAAGTGC 7864
QY 7874 GCGCCACATTCGCGCAAAATCCAAATTTGGCTATCGGGCAAAAGGAGCTCCGGAACCTATC 7933
DB 7865 GCGCCACATTCGCGCAAAATCCAAATTTGGCTATCGGGCAAAAGGAGCTCCGGAACCTATC 7924
QY 7934 CAGGAGGCGCTTAACCAATCCGCTCGTGGGAGGACTTCTGGGAAGACACTGAAC 7993
DB 7925 CAGGAGGCGCTTAACCAATCCGCTCGTGGGAGGACTTCTGGGAAGACACTGAAC 7984
QY 7994 ACCAATTGACACACCATCATCGCCTTATCGTATTTCCAGACCTGGAGTTCTGGTCCACACAGAGAGG 8053
DB 7985 ACCAATTGACACACCATCATCGCCTTATCGTATTTCCAGACCTGGAGTTCTGGTCCACACAGAGAGG 8044
QY 8054 AGCCGCAAGCAGCTCGCCTTATCGTATTTCCAGACCTGGAGTTCTGGTCCACACAGAGAGG 8113
DB 8045 AGCCGCAAGCAGCTCGCCTTATCGTATTTCCAGACCTGGAGTTCTGGTCCACACAGAGAGG 8104
QY 8114 GATGAGCCTTTAGGAGTGGTCTCACCTCTCTCGAGCGCTGATGGGCTCTCTATCGG 8173
DB 8105 GATGAGCCTTTAGGAGTGGTCTCACCTCTCTCGAGCGCTGATGGGCTCTCTATCGG 8164
QY 8174 ATTTCATTAATCTCCCAAGCAGCGGTTCAGTTCTGGTGAATACCTGGAAATCAAAGAA 8233
DB 8165 ATTTCATTAATCTCCCAAGCAGCGGTTCAGTTCTGGTGAATACCTGGAAATCAAAGAA 8224
QY 8234 ATGCCCTATGGCCTTCTCATATGACACCGCGCTTTTGAATCAAGGCTCACTGAGAGTGA 8293
DB 8225 AAACCCCATGGCTTTTCATATGACACTCGCTGTTTCGACTCAACGGTCAACGAGACGA 8284
QY 8294 CATTCGTGTGAGGAGTCAATTTACAGATTTGTGACTTGGCCCGCGAGCGCCAGACAGGC 8353
DB 8285 CATCCGTGTGAGGAGTCAATTTACAGATTTGTGACTTGGCCCGCGAGCGCCAGACAGGC 8344
QY 8354 CATAAGGTGCTCAGACAGCGCTTTACATCGGGGTCCCTGACTAATCAAAAGGSCA 8413
DB 8345 CATAAAATCGCTCAGACAGCGCTTTATATCGGGGTCTCTGACTTAATCAAAAGGSCA 8404
QY 8414 GAATCGGGTTATCGCGGTGCGCGCAAGTGGCTGCTGACACTAGCTGCGGTAAATAC 8473
DB 8405 GAACCTGGGTATCGCGGTGCGCGCGAGCGCGCTGCTGACACTAGCTGCGGTAAACAC 8464
QY 8474 CTTCAATGTTACTTTGAGGCACTGCACCTGCTGAGCTGCAAGCTCCAGACTGGAC 8533
DB 8465 CTTCAATGTTACTTTGAGGCACTGCACCTGCTGAGCTGCAAGCTCCAGACTGGAC 8524
QY 8534 GATGCTGTGAACGAGAGCACTTGTGTATCTGTGAAAGCGCGGGAACCCAGAGGA 8593
DB 8525 GATGCTGTGAACGAGAGCACTTGTGTATCTGTGAAAGCGCGGGAACCCAGAGGA 8584
QY 8594 TCGCGCGGCTTACGAGCTTCAAGAGCTTAAGCTAGGTATTCGCGCCCGCCCGGGGA 8653
DB 8585 TCGCGCGGCTTACGAGCTTCAAGAGCTTAAGCTAGGTATTCGCGCCCGCCCGGGGA 8644
QY 8654 TCGCGCCCAACAGAAATAGCACTGGAGCTGATACATCAATGTTCTCAATGTTGAGT 8713
DB 8645 CCGCGCCCAACAGAAATAGCACTGGAGCTGATACATCAATGTTCTCAATGTTGAGT 8704
QY 8714 CCGCAGCATGATCTGGCAAAAGGTATACCTACCTCAGCGGTGAGCCACACCCCGCT 8773
DB 8705 CCGCAGCATGATCAGCAAAAGGTATACCTACCTCAGCGGTGATCCCGCTGATCCACACCCCGCT 8764

QY 8774 TGCAGCGCTCGTGGGAGACAGCTAGACACACTCCAAATCAACTCTTGGCTAGGCAATAT 8833
DB 8765 AGCAGCGCTCGTGGGAGACAGCTAGACACACTCCAAATCAACTCTTGGCTAGGCAATAT 8824
QY 8834 CATCATGATCGGCCACCCCTATCGGCAAGGATGATCTCTGATGACTCATTTTCTCCAT 8893
DB 8825 TATTATGATCGGCCACTTTTGTGGCAAGGATGATCTCTGATGACTCATTTTCTCCAT 8884
QY 8894 CTTCTAGCTCAAGAGCAACTTGAAGAGCCCTGGATGTCAGATCTACGGGCTTGCTA 8953
DB 8885 CTTCTAGCGAGGAGCAACTTGAAGAGCCCTGGATGTCAGATCTACGGGCTTGCTA 8944
QY 8954 CTCATTTAGCCACTTTGACTTACCTCAGATCAATCAAGCACTCCATGCTTAGCGCAT 9013
DB 8945 CTCATTTAGCCACTTTGACTTACCTCAGATCAATCAAGCACTCCATGCTTAGCGCAT 9004
QY 9014 TACACTCCACAGTCTCTCCAGGTGAGATCAATAGGTTGGCTTCATGCTCAGGAACT 9073
DB 9005 TTCACTCCATAGTTACTCTCCAGGTGAGATCAATAGGTTGGCTTCATGCTCAGGAACT 9064
QY 9074 TGGGTACCCCTTGGCAACTCGAGACATCGGCGCAAGAGTCTCCGCTAGCTACT 9133
DB 9065 TGGGTACCCCTTGGCAACTCGAGTCTGGAGACATCGGCGCAAGAGTCTCCGCTAGCTACT 9124
QY 9134 GTCCAGGGGGGAGGGCGCCACTTTGTGGCAGATACCTTTTAACTGGGCGAGTAAGGAC 9193
DB 9125 GTCCAGGGGGGAGGGCGCCACTTTGTGGCAATACCTTTCACTGGGCGAGTAAGGAC 9184
QY 9194 CAGCTTAACTCACTCCATTCGCGCGCTCCAGCTGAGCTGAGTGTGCTGCTGCT 9253
DB 9185 CAAACTTAACTCACTCCATTCGCGCGCTCCAGCTGAGCTGAGTGTGCTGCTGCT 9244
QY 9254 CGCTGTACAGCGGGGAGACATATATCAGAGCTGCTCTCGTCCGAGCCCGCTGCT 9313
DB 9245 TGCTGTTACAGCGGGGAGACATATATCAGAGCTGCTCTCGTCCGAGCCCGCTGCT 9304
QY 9314 TCCGTTGTGCTACTCTCTACTTTCTGTAGGGTAGGCAATTTACCTGCTCCCAACCGATG 9373
DB 9305 CATGCTGTGCTACTCTCTACTTTCTGTAGGGTAGGCAATTTACCTGCTCCCAACCGATG 9364
QY 9374 AACGGGAGCTAACCATCCAGGCTTAAGCCATTTCTGTTTTTTTTTTTTT 9425
DB 9365 AACGGGAGATAAACAATCCAGGCAATAGGCCATCCCTTTTTTTTTTTTTT 9416

RESULT 4
US-08-904-686A-1
; Sequence 1, Application US/08904686A
; Patent No. 5998130
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLan &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,686A

QY	9194	CAAGCTTAACACTCACTCCAAATCCGGCCGGGTCCTCCAGCTGGAGCTGTGCTGGCTGGTTTCGT	9255
Db	9185	CAAACTTAAACTCACTCCAAATCCGGCTCGTCCCGCTGGACTGTGTCGGCTGGTTTCGT	9244
QY	9254	CGCTGGTTACAGCGGGGAGACATATATCACAGCCTGTCTCGGTGCCCGACCCCGCTGGTT	9313
Db	9245	TGCTGGTTACAGCGGGGAGACATATATCACAGCCTGTCTCGTGCCTCCCGCTGGTT	9304
QY	9314	TCCGTTGTGCTTACTCCTTACTTCTGTAGGGTACGATTTACCTGCTCCCAACCGATG	9373
Db	9305	CATGCTGTGCTTACTCCTTACTTCTGTAGGGTACGATTTACCTGCTCCCAACCGATG	9364
QY	9374	AACGGGAGTACCACTCCAGCGCTTACGCAATTTCCCTGTTTTTTTTTTTT	9425
Db	9365	AACGGGAGTAAACACTCCAGGCAATAGGCCATCCCCCTTTTTTTTTT	9416

```

RESULT 5
US-09-315-850-1
; Sequence 1, Application US/09315850
; Patent No. 6217872
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLend &
; SURFET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,850
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,686
; FILING DATE: 01-AUG-1997
; APPLICATION NUMBER: US 08/324,977
; FILING DATE: 18-OCT-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McLeLend, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 900703G

```


QY 1934 GAATGAGACAGACGTGATGCTCTCTCAACAAACAGCGTCCGCCCAAGGCAAACTGGTTGG 1993
Db 1925 GAAACGAGACTGACGTGCTGCTGCTCAACAAACAGCGCGCGCAAGCAACAGCTGGTTCGG 1984
QY 1994 CTGTACATGGAATGATAGTACTGGTTTCACTAAGACGTGCGAGGFTCCCCCGTGTATCAT 2053
Db 1995 CTGCATGGAATGAAATAGACACCGGTTTCAACAAACATGTGGGGGCCCCCGTGTAACT 2044
QY 2054 CGGGGGGTCTGGTAACCCCACTTGATCTGCCCAACAGCACTGCTTCCGGAAGACACCCGA 2113
Db 2045 CGGGGGGTTCGCAACACACACCTTGACCTGCCCAACAGCACTGCTTCCGGAAGACACCCGA 2104
QY 2114 GGCTACTTACACAAATGTGCTGGGCCCTGTGTGACACTAGGTGCTTAGTACTA 2173
Db 2105 GGCTACTTACACAAATGTGCTGGGGCTTGGCTGACACTAGGTGCTTAGTACTA 2164
QY 2174 CCCATACAGGCTTGGCACTACCCCTGACCTTCAATGCGCATGCAATGGACTCGAGGAGAGCG 2233
Db 2165 TCCATACAGGCTTGGCACTACCCCTGACCTTCAATGCGCATGCAATGGACTCGAGGAGAGCG 2224
QY 2234 GTATGTGGGGGTGGAGCACAGGCTCAATGCGCATGCAATGGACTCGAGGAGAGCG 2293
Db 2225 GTATGTGGGGGTGGAGCACAGGCTCAATGCGCATGCAATGGACTCGAGGAGAGCG 2284
QY 2294 CTGTAACTTGGAGACAGGATAGGTGAGAACTCAGCCCGCTGTGTGTCTACAAAGA 2353
Db 2285 TTGTGACTTGGAGACAGGATAGGTGAGAACTCAGCCCGCTGTGTGTCTACAAAGA 2344
QY 2354 GTGCGAGACTGCGCTGTGCTTCAACACACCTACCGCTTATCCACTGGTTGATCCA 2413
Db 2345 GTGCGAGTACTGCGCTGTGCTTCAACACACCTACCGCTTATCCACTGGTTGATCCA 2404
QY 2414 TCTCCATCAGAACATGCTGGACGTGCAATACCTGTACGGTGTAGGGTCAAGGTTGTCTC 2473
Db 2405 CCTCCATCAGAACATGCTGGACGTGCAATACCTGTACGGTGTAGGGTCAAGGTTGTCTC 2464
QY 2474 CTTTGCATCAAAATGGAGTACATCTCTGTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2533
Db 2465 CTTTGCATCAAAATGGAGTATGCTCTGTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2524
QY 2534 GTGTGCTGCTGTGGATGATGCTGTGATAGCCAGGCTGAGCCGCCCTTATAGAACTT 2593
Db 2525 CTGTGCTGCTGTGGATGATGCTGTGATAGCCAGGCTGAGCCGCCCTTATAGAACTT 2584
QY 2594 GTGTGCTCTCAATGCGGCTCGTGGCCGAGCGCATGGTATTCTCTCTCTCTCTCTCTCTCT 2653
Db 2585 GTGTGCTCTCAATGCGGCTCGTGGCCGAGCATGGCATCTCTCTCTCTCTCTCTCTCTCTCT 2644
QY 2654 CTTCTGCGCGCTGTGATACATTAAGGAGAGGCTGGCTCTCTGGGCGCGGTATGCTTTTAA 2713
Db 2645 CTTCTGCGCGCTGTGATACATTAAGGAGAGGCTGGCTCTCTGGGCGCGGTATGCTTTTAA 2704
QY 2714 TGGCGTATGCGCGCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2773
Db 2705 TGGCGTATGCGCGCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2764
QY 2774 CCGGAGATGCTGATCTGTGCGGGGTGCGGTTCTGTGTAGGTTGTGTTTCTCTCTCTCT 2833
Db 2765 CCGGAGATGCTGATCTGTGCGGAGCGGTTTGTGGTCTGTGTTCTCTCTCTCTCTCTCT 2824
QY 2834 GTCAACATCTACAAAGTGTCT 2893
Db 2825 GTCAACATCTACAAAGTGTCT 2884
QY 2894 CAGAGCGAGCGGACATGCAAGTGTGGGTCCCCCGCTCAACGTTTCGGGAGCGCGCGA 2953
Db 2885 CAGAGCGAGCGGAGCTACATGTGTGATGCCCGCTCAACGTTTCGGGAGCGCGCGA 2944
QY 2954 TGCCATCATCT 3013
Db 2945 TGCCATCATCT 3004

QY 3014 CCTGCTGCCATCTCGGCCGCTCATGCTGTCTCCAGCTTGGCAATAACGAGAGTCCGTA 3073
Db 3005 TCATATGTCATCTCGTTCGCTCATGCTGTCTCCAGCTTGGCAATAACGAGAGTCCGTA 3064
QY 3074 CTTCTGTCGCGCTCAAGGGCTCATTTCTGTCATGTCATGTTAGTGCAGAAATGTCGCGGGG 3133
Db 3065 CTTCTGTCGCGCTCAAGGGCTCATTTCTGTCATGTCATGTTAGTGCAGAAATGTCGCGGG 3124
QY 3134 TCATATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 3193
Db 3125 TCATATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 3184
QY 3194 CCATCTTACCCCTCGGACTTGGCCCAACGCGGCTTACGAGACTTTCGCGTGGCGGT 3253
Db 3185 CCATCTTACCCCTCGGACTTGGCCCAACGCGGCTTACGAGACTTTCGCGTGGCGGT 3244
QY 3254 AGAGCCCGTCTCTTCTCCGCAATGAGACCAAGTCTATCCTTGGGAGACAGACCCG 3313
Db 3245 GGAGCCCGTCTCTTCTCCGCAATGAGACCAAGTCTATCCTTGGGAGACAGACCCG 3304
QY 3314 TCGTGTGGGACATCATCTTGGTCTTACCCGCTCCGCGGAGGAGGAGATATT 3373
Db 3305 GGCTGTGGGACATCATCTTGGTCTTACCCGCTCCGCGGAGGAGGAGATATT 3364
QY 3374 TTTGGACCGGCTGATAGTCTCGAAGGCAAGGCTGGGACTCTCTTCCGCCCATCACGGC 3433
Db 3365 CTTGGGCGCGGCTGATAGTCTTGAAGGCGGGGTTTGGACTCTCTCGCGCCCATCACGGC 3424
QY 3434 CTACTTCCCAACAAACGCGGGGCTACTTGGTTCATCATCATCTAGCTCATAGGCGGGGA 3493
Db 3425 CTACTTCCCAACAAACGCGGGGCTACTTGGTTCATCATCATCTAGCTCATAGGCGGGGA 3484
QY 3494 CAAGAACAGGTCGAGGAGGTTTCAAGTGGTTCCTACCGCAACACAACTTCTCTCTGGC 3553
Db 3485 CAGAACAGGTCGAGGAGGTTTCAAGTGGTTCCTACCGCAACACAACTTCTCTCTGGC 3544
QY 3554 GACTGCTCAACGCGGCTGTGTGACTGTCTACCATGCGCTGTCTCGAAGACCTTAGC 3613
Db 3545 GACTGCTCAACGCGGCTGTGTGACTGTCTACCATGCGCTGTCTCGAAGACCTTAGC 3604
QY 3614 CGTTCGCAAGGTCATCATCCCAATGTACCAATGTAGACTCTGCACTCTGCTGCGGTG 3673
Db 3605 CGCGCAAGGAGGCTCAATCACCCAGATGTACATAATGTGGACCAAGCACTCTCTGCGGTG 3664
QY 3674 GCAAGCGCGCGCGCGCTTCCATGACACCATGCTGTGGAGCTTCAAGACCTTTA 3733
Db 3665 GCCCAAGCGCGCGCGCTTCCATGACACCAATGTAGACTCTGCACTCTGCTGCGGTG 3724
QY 3734 CTTGGTCAAGAGACATGCTGATGTCTATTCCGCTGCGCGCGGCGAGCGACAGCAGGCGAAG 3793
Db 3725 CTTGGTCAAGAGACATGCTGATGTCTATTCCGCTGCGCGCGGCGAGCGACAGTAGGCGGAG 3784
QY 3794 TCTACTCTCCCGAGCGCTCTCTACTCTGAAAGCTCTCTGGGTGGTCCATTTGTTG 3853
Db 3785 CTTGCTCTCTCCCGAGCGCTCTCTACTTTGAAGGCTCTTGGGTGGTTCACCTGCTCTG 3844
QY 3854 CTTCTGGGCGAGCTGCTGGGCGCTCTTCCGCGGCTGTGTGTGACACCGCGGGGCTCCGAA 3913
Db 3845 CTTCTGGGCGAGCTGCTGGGCGCTCTTCCGCGGCTGTGTGTGACACCGCGGGGCTTCGAA 3904
QY 3914 GCGGTGGGACTTCACTACCGGTGAGTCTATGAAACTACCATGCGGTCTCCGCTCTTAC 3973
Db 3905 GCGGTGGGACTTGTGTCGCGGTAGAGTCCATGGAATCTATGCGGTCTCTCGGTCTTAC 3964
QY 3974 AGACAACTCAACCCCGCTGTACCGCAGACATTTCCBAAGTGGCACATCTCACGCTCC 4033
Db 3965 GGAACACTCATCCCCCGCGCTTACCGAGTCAITTTAAGTGGGCCACTTACGCTCC 4024
QY 4034 TATGTGACGCGCAGAGACCAAAAGTGGCGCTGCGTATGAGGCCCAAGGTTAAGGT 4093
Db 4025 CACTGCGCGGCAAGAGTACTTAAAGTGGCGCTGCATATGACGCCCAAGGTTAAGGT 4084
QY 4094 GCTGCTCTGAAACCGCTGCTGCGCGCACCTTTAGGGTTTGGGCGCTATATGTTCCAAAGC 4153

Db	4085	GCCTGCTCTCAATCCGTCGCTTCCGCTTACCTTAGGGTTTGGGCGATATATGCTTAAGGC	4144
Qy	4154	ACACGGTATCGACCCCTACATCAGAACTGGGGTAAAGACCATTAACACAGCGCGCTCCAT	4213
Db	4145	ACACGGTATCGACCCACATCAGAACTGGGGTAAAGACCATTAACACAGCGCGCCCGT	4204
Qy	4214	TACGTACTCCACCTATAGGCATGTCCTTTGCGCAGCGTGGCTGTCTCTGGGGCGCTATGA	4273
Db	4205	CACATACTCTACCTATGCAAGTTCTTGCCGATGGTGTGTGCTCTGGGGCGCTATGA	4264
Qy	4274	CATCATATATGTCATAGTGCCTCAACTGACTCGACTACCACTTTGGGCATPGGCAC	4333
Db	4265	CATCATATATGTCATAGTGCCTCAACTGACTCGACTACCACTTTGGGCATPGGCAC	4324
Qy	4334	AGTCCTGGACCAAGCGGAGACGGCTGGAGCGCGCTCGTCTGCTGCCACCGCTACAC	4393
Db	4325	AGTCCTGGACCAAGCGGAGACGGCTGGAGCGCGCTGTGTCTGCTGCCACCGCTACGCC	4384
Qy	4394	TCCGGGATCGTTACCTGTGCACACCCCAATATCGAGGAAATAGGCTGTCCACCAATGG	4453
Db	4385	TCCGGGATCGTCAACCTGTGCACACCCCAACATCGAGAGGTGCGCTGTCTAATCTGG	4444
Qy	4454	AGAGATCCCTTCTATGGCAAGCCATCCCATTTAGAGCCATCAGGGGGGGAAGGCATCT	4513
Db	4445	AGAGATCCCTTCTATGGCAAGCCATCCCATTTAGAGCCATCAGGGGGGGAAGGCATCT	4504
Qy	4514	CATTCTTGCCATTTCCAAAGAGAAATGTGACGAGCTCGCGCGAAAGCTGACAGGCCTCG	4573
Db	4505	CATTCTTGCTCATTTCCAAAGAGAAAGTTCGACGAGCTCGCGCGAAAGCTGACAGGCCTCG	4564
Qy	4574	ACTGAACGCTGTAGCATATTAACGGGGCTTTGATGTTCGCTCATACCGCCTATCGGAGA	4633
Db	4565	AATCAACGCCTGTGGCGTATTAACGGGGGCTCGATGTGTCCGTATACCAACTATCGAGA	4624
Qy	4634	CGTCGTTGCTGTGGCAACAGACGCTCTGATGACGGGCTATACGGGGGACCTTTGACTCAGT	4693
Db	4625	CGTCGTTGCTGTGGCAACAGACGCTCTGATGACGGGCTATACGGGGGACCTTTGACTCAGT	4684
Qy	4694	GATCGACTGCAATACATGTGCTACCCAGACAGTCGACTTCAGCTTCGGATCCCACTTCAC	4753
Db	4685	GATCGACTGTAAACATGTGTACCCACAGACAGTCGACTTCAGCTTCGGATCCCACTTCAC	4744
Qy	4754	CATTGACAGACACGCGTGCCCAAGACGCGGTGTCCGCTCGCAACGCGGAGGTAGAAC	4813
Db	4745	CATTGACAGACACGCGTGCTCAAGACGCAAGTGTCCGCTCGCAACGCGGAGGTAGAAC	4804
Qy	4814	TGCGAGGGTAGAGTGGCATCTACAGGTTTGTGACTCCAGAGAACGCGCCCTCGGCGAT	4873
Db	4805	TGCGAGGGTAGAGAGGCACTACAGGTTTGTGACTCCCGGAGAACGCGCCCTCGGCGAT	4864
Qy	4874	GTTTCGATTCCTCGGTCTCTGTGTGAGTGCCTATGACGCGGCTGTGTGTGTATGAGCTCAC	4933
Db	4865	GTTTCGATTCCTCGGTCTCTGTGTGAGTGCCTATGACGCGGCTGTGTGTGTATGAGCTCAC	4924
Qy	4934	GCCCGCTGAGACCTCGGTTAGGTTGGGGCTTACCTAAATACACAGGGTTGCCCGCTTG	4993
Db	4925	CCCGCCCGAGACCTCGGTTAGGTTGGGGGCTTACCTAAATACACAGGGTTGCCCGCTTG	4984
Qy	4994	CCAGGACCACTCTGGAGTCTGGGAGAGCGTCTTCAGAGCCTCACCCCATATGATGCCCA	5053
Db	4985	CCAGGACCACTCTGGAGTCTGGGAGAGTGTCTTCAGAGCCTCACCCCATATGATGCCCA	5044
Qy	5054	CTTCTGTCTCCAGACTAAACAGGCAGGAGCAACTTTCCTTACCTGTGGCATNTCAAGC	5113
Db	5045	CTTCTGTCTCCAGACTAAACAGGCAGGAGCAACTTTCCTTACCTGTGGCATNTCAAGC	5104
Qy	5114	TACAGTGTGCGCAGGGCTCAAGCTCCACTCGATGTTGGGACCAAAATGTGGAAGTGTCT	5173
Db	5105	CACGGTGTGCGCAGGGCTCAGSCCCCACTTCCATCTGGGATCAATGTGGAAGTGTCT	5164
Qy	5174	CATACGGCTGAACCTTACACTCCAACGGGCGCAACCCCTGCTGTATAGCTAGGACCGT	5233

Db	5165	CATACGGGTGAAACCTACGCTGCACGGGGCCAAACCCCTTGCTGTACAGGCTGGGAGCGGT	5224
Qy	5234	CCAAATGAGGTGTCCTCTCACACACCCCAATAAATACATATGCATGCAATGCATGCGG	5293
Db	5225	CCAGATGAGGTGACCCCTCACCCACCCCAATAACCAATATACATATGCAATGCAATGTCGCG	5284
Qy	5294	TGACCTGAGGTGCTGCTACTAGCACCTGGGTCTGTGTAGGCGGAGTCTCTGCAGCTTTGGC	5353
Db	5285	TGACCTGAGGTGCTGCTACTAGCACCTGGGTGCTGTGTGGCGGAGTCTCTGCAGCTCTGGC	5344
Qy	5354	CGCATACTGCTGTAGGACAGACGACGTGTGTCTATGTGGCGAGGATCATCTTTGTCCGGGAA	5413
Db	5345	CGCGTATTGCTGTAGCAACAGGCACTGGTTCATTTGTGGGTAGGATTAATCTTTGTCCCGGAG	5404
Qy	5414	GCCAGCTGTGCTTCCCGACAGGGAACTCTCTACAGGAGTTGATGAGATGAAGAGTG	5473
Db	5405	GCCGGCCATTGTTCCCGACAGGAGCTTCTCTACAGGAGTTGATGAATGAAGAGTG	5464
Qy	5474	TGCTTCACAACTTCTCTTACATCGAGCAGGGAATGCAGTCCCGAGCAATTCAGCAAAA	5533
Db	5465	CGCCTCGCACCTCTTACATCGAGCAGGGAATGCAGTCCCGAGCAATTCAGCAGAA	5524
Qy	5534	GCGCTCGGGTGTGTGAAGCGGCACCAAGCAGGAGGCTGTCTCTCCCTGTGTGGA	5593
Db	5525	AGCCTCGGGTACTGTCAAAACAGCCAAACAGCGGAGGCTGCTCTCCCTGTGTGGA	5584
Qy	5594	GTCCAAGTGGCGAGCCCTTGAGACCTTCTCGGCGAAGCACATGTGAAATTCATCAGCGG	5653
Db	5585	GTCCAAGTGGCGAGCCCTTGAGACATTTCTGGGCGAAGCACATGTGAAATTCATCAGCGG	5644
Qy	5654	AATTCAGTACTACAGGCTTATCCACTGTCCCTTGAAACCCCGCATAGCATCATTTGAT	5713
Db	5645	GATACAGTACTTACAGAGCTTATCCACTGTCCCTTGGAACCCCGCAATAGCATCATTTGAT	5704
Qy	5714	GGCATTTACAGCTTCTATCATAGCCGCTCACACCCAAACACCCCTCTCTGTTTAACT	5773
Db	5705	GGCATTTACAGCTTCTATCATCCAGCCGCTTCAACCCAAAGTACGCTCTCTGTTTAACT	5764
Qy	5774	CTTGGGGGATGGTGGCTGCCAACTCGCTCTCCAGCGTCCGCTGAGCTTTCTGTTGGG	5833
Db	5765	CTTGGGGGGTGGTGGCTGCCAACTCGCCCCCCCCAGCGCGCTTCGCGCTTCTGTTGGG	5824
Qy	5834	CGCGGGCATCGCCGAGCGGCTTTGGCAGCATAGGCTTGGAAGTGTCTCTGAGCAT	5893
Db	5825	CGCGGGCATCGCGGTGGGCTGTGGCAGCATAGGCTTGGAAGTGTCTGTTGACAT	5884
Qy	5894	CTTGGCGGGTATGGGCGAGGGTAGCGGGCACTCTGTGGCTTTTAAAGTCATGAGCGG	5953
Db	5885	TCTGGCGGGTATTGAGCAGGAGTGGCGCGCGGCTCTGTGGCTTTTAAAGTCATGAGCGG	5944
Qy	5954	CGAGTGTCCCTCCACCGAGGACCTGTGCAACTTACTCCCTGCCATCTCTCTCTCTGTTGC	6013
Db	5945	CGAGATGCCCTCCACCGAGGACCTGTGCAACTTACTCTCTGCCATCTCTCTCTCTGTTGC	6004
Qy	6014	CTTGGTCTGTGGGCTCTGTGGCAGCAATACTGTGGTTCGACATAGGCTTGGAAGTGTCTCTGAGGAGGG	6073
Db	6005	CCTGGTCTGTGGGCTCTGTGTGTGACAAATACTGTGGTTCGACATAGGCTTGGAAGTGTCTCTGAGGAGGG	6064
Qy	6074	GGCTGTGCAGTGATGAACCGGCTGTATGAGCTTGTGGCTTTCGCGGGGTAAACAGTCTGCC	6133
Db	6065	GGCTGTGCAGTGATGAACCGGCTGTATGAGCTTGTGGCTTTCGCGGGGTAAATCATGTTTCCC	6124
Qy	6134	TAGCCTACTGTCCCTGAGAGGACGCTCGACAGCTGTCATCTAGATCTCTCTAGCT	6193
Db	6125	CAGCCTATGTCCCTGAGAGGACGCGCAGCGGCTGTATCTCAGATCTCTCTCAGGCT	6184
Qy	6194	TACCATCACTCACTGCTGAAGCGGCTCCACAGTGGATTAAATGAGGACTGTCTTACGCC	6253
Db	6185	TACCATCACTCACTGCTGAAGGCTCCACAGTGGATTAAATGAGGACTGTCTCACC	6244
Qy	6254	ATGCTCGGGCTGTGCTTAAGGAGTGTGGGATTTGGATATGCACGGTGTGACTGACTT	6313
Db	6245	GTCTTCGGGCTGTGCTTAAGGAGTGTGGGACTGGATATGCACGGTGTGACTGACTT	6304

QY 6314 CAAGACCTGGCTCAGTCCAAAATCTCTGCGGCGGTACCGGAGTCCCTTTCTTGTCATG 6373
DB |||||
QY 6305 CAAGACCTGGCTCAGTCCAAAATCTCTGCGGCGGTACCGGAGTCCCTTTCTTGTCATG 6364
DB |||||
QY 6374 CCAACGCGGTTACAAGGAGTCTGGCGGGGAGCGGCATCATCAAAACCACTGCCCATG 6433
DB |||||
QY 6365 CCAACGCGGTTACAGGAGTCTGGCGGGGAGCGGCATCATCAAAACCACTGCCCATG 6424
DB |||||
QY 6434 CGGAGCACAGATCCCGGACATGTCAAAAACGGTTCCATGAGGATCGTAGGGCTTAGAAC 6493
DB |||||
QY 6425 TGGAGCACAGATCACCGGACATGTCAAAAACGGTTCCATGAGGATCGTAGGGCTTAGAAC 6484
DB |||||
QY 6494 CTCGAGCAACAGTGGACCGGAGTCTCCCAATCAACGATACACGAGGAGTGGGTCAC 6553
DB |||||
QY 6495 CTCGAGCAACAGTGGATCGGAGATTCGCCATCAACGATACACGAGGAGTGGGTCAC 6544
DB |||||
QY 6554 ACCCTCCCGCGCGCCAACTATTCCAGGCGCTATGGCGGGTGGCTGCTGAGGAGTACGT 6613
DB |||||
QY 6545 ACCCTCTCCAGCGCCAACTATTCTAGGCGCTGTGGCGGGTGGCGCTGAGGAGTACGT 6604
DB |||||
QY 6614 GGAGGTACCGGTGGGGATTCACACTACGTACCGGCGATGACCACTGACAACTGATAA 6673
DB |||||
QY 6605 GGAGGTACCGGTGGGGATTCACACTACGTACCGGCGATGACCACTGACAACTGATAA 6664
DB |||||
QY 6674 GTGCCATGCCAGTTCGCGCCCGGAACTTCTTCAAGGAGTGGAGTGGGTTGCA 6733
DB |||||
QY 6665 GTGCCATGCCAGTTCGCGCCCGGAACTTCTTCAAGGAGTGGAGTGGGTTGCA 6724
DB |||||
QY 6734 CAGGTACGCTCCCGCGTGCAAAACCTCTTCAAGGAGGAGCTCACGTTCAGAGTCCGGCT 6793
DB |||||
QY 6725 CAGGTACGCTCCCGCGTGCAGAGCTCTCTTCAAGGAGGAGTTCATTCAGAGTCCGGCT 6784
DB |||||
QY 6794 CAACCAATACCTGGTGGGTTCAGAGTACCATGCGAGCGGAGTGGGTTGCA 6853
DB |||||
QY 6785 CAACCAATACCTGGTGGGTTCAGAGTACCATGCGAGCGGAGTGGGTTGCA 6844
DB |||||
QY 6854 TACTTCCATGCTCACCGATCCCTCCACATATACAGAGAGACGCTAACGTTAGGCTGGC 6913
DB |||||
QY 6845 CATTTCATGCTCACCGACCCCTCCACATACAGAGAGAAACGCTAACGTTAGGTTGGC 6904
DB |||||
QY 6914 TAGAGGTCTCCCGCTCTTTAGCCAGCTCATCAGCTAGCCAGTCTCTGCGCTCTCTTT 6973
DB |||||
QY 6905 CAGGGGTCTCCCGCTCTCTGCGAGCTCTCAGTACCCAGTCTCTGCGCTCTCTTT 6964
DB |||||
QY 6974 GAAGGAGCATGACATACCCACCATGCTCTCCGAGCGCTGACCTCATCGAGGCCAACCT 7033
DB |||||
QY 6965 GAAGGAGCATGACATACCCACCATGCTCTCCGAGCGCTGACCTCATCGAGGCCAACCT 7024
DB |||||
QY 7034 CTGTGGCGGAGGAGTGGGCGGAAACATCACTCCGTTGGAGTCAAGATTAAGTACT 7093
DB |||||
QY 7025 CTTGTGGCGGAGGAGTGGGCGGAAACATCACTCCGTTGGAGTCAAGATTAAGTACT 7084
DB |||||
QY 7094 AATTCTGGACTCTTTCCGAACCGCTTCAAGGAGGAGGAGTCAAGGAGTCAAGGAGT 7153
DB |||||
QY 7085 AGTCTCGGACTCTTTCCGACCCGCTTCAAGGAGGAGGAGTCAAGGAGTCAAGGAGT 7144
DB |||||
QY 7154 GCGGAGTCTCGGAAATCAAGGAGTTCCTCAGGTTGGCCATTAAGGAGTCAAGGAGT 7213
DB |||||
QY 7145 GCGGAGTCTCGGAAATCAAGGAGTTCCTCAGGTTGGCCATTAAGGAGTCAAGGAGT 7204
DB |||||
QY 7214 GGACTACAATCTCCACTGCTAGGTCCTGAGGAGGAGGAGTCAAGGAGTCAAGGAGT 7273
DB |||||
QY 7205 GAGTACACACCTCCACTGTTAGAGTCTTGGAGGAGGAGTCAAGGAGTCAAGGAGT 7264
DB |||||
QY 7274 ACACGATGCCATTCGCCACCTACCAAGGCTCTCCATATACACCTCCACGAGGAGGAG 7333
DB |||||
QY 7265 GACAGGTCCTGTCACGAGTCAAGGAGTTCCTCAGGAGTCAAGGAGTCAAGGAGT 7324
DB |||||
QY 7334 GAGGTTCTGTCAGAGATCAAGTGTCTGCTTGGGAGTCTGCGGAGTCAAGGAGT 7393
DB |||||
QY 7325 GAGGTTCTGTCAGAGATCAAGTGTCTGCTTGGGAGTCTGCGGAGTCAAGGAGT 7384
DB |||||

QY 7394 CTTTCGTAGCTCCGGATCTCGGCTCGCTTATAGCGGACCGGAGTCCCTTCTTGACCT 7453
DB |||||
QY 7385 CTTTCGTAGCTCCGGATCTCGGCTCGCTTATAGCGGACCGGAGTCCCTTCTTGACCT 7444
DB |||||
QY 7454 GGCCTTCGAGAGCGGTGACAAAGGATCCGAGTTCGAGTTCGAGTTCCTTCTTGAGTTCGAGT 7513
DB |||||
QY 7445 GGCCTTCGAGAGCGGTGACAAAGGATCCGAGTTCGAGTTCGAGTTCCTTCTTGAGTTCGAGT 7504
DB |||||
QY 7514 TGAAGGGGAGCGGGGAGCGGATCTAGCGGACCGGCTCTTGAGTTCGAGTTCGAGTTCGAGT 7573
DB |||||
QY 7505 TGAAGGGGAGCGGGGAGCGGATCTAGCGGACCGGCTCTTGAGTTCGAGTTCGAGTTCGAGT 7564
DB |||||
QY 7574 GGTAGTTCGAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7633
DB |||||
QY 7565 AGCTAGTTCGAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7624
DB |||||
QY 7634 GCATGCGCTCGGAGGAGGAGTAAAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7693
DB |||||
QY 7625 GCATGCGCTCGGAGGAGGAGTAAAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7684
DB |||||
QY 7694 TCACCAACATGCTCTACGCGCAACATCCGCGAGCGGCAAGCTCCGCGAGGAGGAGT 7753
DB |||||
QY 7685 CCACCAATACATGCTTATGCGCAACATCTCGAGCGGAGGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7744
DB |||||
QY 7754 CACCTTTGAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7813
DB |||||
QY 7745 CACCTTTGAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7804
DB |||||
QY 7814 GCGGAGGCTCCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7873
DB |||||
QY 7805 GCGGAGGCTCCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7864
DB |||||
QY 7874 GCGGAGGCTCCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7933
DB |||||
QY 7865 GCGGAGGCTCCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7924
DB |||||
QY 7934 GAGGAGGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7993
DB |||||
QY 7925 CAGCAAGGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 7984
DB |||||
QY 7994 ACCAATTCAGACCCAGCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8053
DB |||||
QY 7985 ACCAATTCAGACCCAGCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8044
DB |||||
QY 8054 AGGCGGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8113
DB |||||
QY 8045 AGGCGGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8104
DB |||||
QY 8114 GATGCGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8173
DB |||||
QY 8105 GATGCGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8164
DB |||||
QY 8174 ATTTCAATTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8233
DB |||||
QY 8165 ATTTCAATTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8224
DB |||||
QY 8234 ATGCGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8293
DB |||||
QY 8225 AAAACCCATGCGCTTCAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8284
DB |||||
QY 8294 CATTGCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8353
DB |||||
QY 8285 CATCGCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8344
DB |||||
QY 8354 CATAGGTCGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8413
DB |||||
QY 8345 CATAGGTCGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8404
DB |||||
QY 8414 GAACTGCGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8473
DB |||||
QY 8405 GAACTGCGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8464
DB |||||
QY 8474 CCTCACTGTTACTTTGAAGGCTTCAGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 8533
DB |||||

QY	494	GGCTTCCGAGCGGTGCGAACCTGCTGGAGGGCGACAACTATCCAAAGGCTCGCCGACC	555
DB	495	GACTTCCGAGCGGTGCGAACCTGCTGGAGGGCGACAACTATCCCAAGGCTCGCCGACC	544
QY	554	CGAGGGCAGGGCTTGGGCTCAGCCCGGGTACCCCTTGGCCCTCTATGGCAATGAGGGCT	613
DB	545	CGAGGGCAGGACCTTGGGCTCAGCCCGGGTACCCCTTGGCTCTCTATGGCAATGAGGGCTT	604
QY	614	GGGGTGGGAGGATGGCTCCGTGACCCCGGGCTCCCGGGCTAGTTGGGGCCCCACGGA	673
DB	605	AGGGTGGGAGGATGGCTCCGTGACCCCGGGCTCCCGGGCTAGTTGGGGCCCCACGGA	664
QY	674	CCCCGGCGTAGTGGCGCTAACTCTGGGTAAAGGTATCGATACCCCTTACATGGGCTTCGC	733
DB	665	CCCCGGCGTAGTGGCGCTAACTTGGGTAAAGGTATCGATACCCCTTACATGGGCTTCGC	724
QY	734	CGATCTCANGGGTACATTCGCGTCTCGTGGGCGCCCCCTAGGGGGCGTGGCAGGGCCCTT	793
DB	725	CGATCTCATGGGGTACATTCGCGTCTCGTGGGCGCCCCCTAGGGGGCGTGGCAGGGCCCT	784
QY	794	GGCACACGCTGTCGGGTTCGTGGAGGACGGCGTGAACATGCAACAGGGAATTGCCCGG	853
DB	785	GGCACATGGTTCGCGGTTCGTGGAGGACGGCGTGAACATGCAACAGGGAATTGCCCGG	844
QY	854	TTGCTCTTTCTCTATCTCTCTCTTGGCTCTGCTGCTCCGTGTTGACATCCAGCTTCGCG	913
DB	845	TTGCTCTTTTCTATCTCTCTTGGCTCTGCTGCTCCGTGTTGACATCCAGCTTCGCG	904
QY	914	TTATGAAGTGGCAACGCTGTCGGGATATACATGTACGAACGACTGCTCCAACTCAAG	973
DB	905	TTACGAAGTGCACACGCTGTCGGGATATATCACTCAGCAACGACTGCTCCAACTCAAG	964
QY	974	CATTCTGTATGAGCAGCGGACGTGATCATGCATCTCCGGTGGCTGCTGCTGTTCA	1033
DB	965	CATTCTGTATGAGCAGCGGACCTTGTATCATGCATCTCTGGTGGCTGCTGCTGTTCA	1024
QY	1034	GGAGGGTAAACGCTCCCGTCTGCTGGTAGGCTCACTCCACGCTCGCGGCAGGAATGC	1093
DB	1025	GGAGGGCACTCTCTCCGCTGCTGGTAGGCTCACTCCACGCTCGCGGCAGGAATGC	1084
QY	1094	CAGGCTCCCACTACGACAATACGACGCCACGTCGACTGCTGCTGGGACGGTGTGTTT	1153
DB	1085	CACCATCCCCACCGACGATACGACGCCACGTCGACTGCTGCTGGGACGGTGTGTTT	1144
QY	1154	CTGCTCCGCTATGTACGTTGGGGATCTCTGCGGATCTATTTTCTCGTCTCCGAGCTGTT	1213
DB	1145	CTGTTCCGCTATGTACGTTGGGGACCTCTGCGGATCTGTTTTCTCGTCTCTCAGCTGTT	1204
QY	1214	CACCTTCTCGCTCCCGCATGAGACAGTCAGAGCTGCAACTGCTCAATCTATCCCGG	1273
DB	1205	CACCTTCTCGCTCCCGGCATGTGACATTCAGGACTGTAACTGCTCAATTTATCCCGG	1264
QY	1274	CCATGTATCAGGTACCGCATGCTTGGGATATGATGAATGAACTGGTCACTCAACAGC	1333
DB	1265	CCATGTGTGGGTCACCGTATGGCTTGGGACATGATGAATGAACTGGTCACTCAACAGC	1324
QY	1334	CTTATGCTGTCGAGTGTCTCGGATCCGACAGCTGTGCTGACATCTGGTGGCGGGGC	1393
DB	1325	CTTATGCTGTCGAGTGTCTCCCGATCCCAAGCGCTGTGCTGACATCTGGTGGCGGGGC	1384
QY	1394	CCACTGGGAGTCTCGCGGGCTTGCCTACTATTTCATGTTAGGGAACCTGGGCTTAAGGT	1453
DB	1385	CCACTGGGAGTCTCGCGGGCTTGCCTACTATTTCATGTCGGGGAACTGGGCTTAAGGT	1444
QY	1454	TCTGATTTGTGGCGCTACTCTTTTGGCGGCTTGAACGGGGAGACCCACAGCGGGAGGGT	1513
DB	1445	TCTGATTTGATGTACTTTTGTGCGGTTGACGGGATACCCACGTCGACGGGGGGC	1504
QY	1514	GGCGGGCCACACCTTCGGGTTTCAGTCCCTTTTCTACTCTGGGGGCTCTAGAAAT	1573
DB	1505	GCATGCCAAACCCACCAAGGCTGCTGTCCATGTTGCAAGTGGGCGCTCTAGAAAT	1564

Qy	1574	CCAGCTTGTTGAATACCAACGGCAGCTGGACATCAACAGGACTGCCCTAAATTGCAATGA	163
Db	1585	CCAGCTTATAAACCAACCAATGGAGTTGGCACAATCAACAGGACTGCCGTGAATGCAATGA	1624
Qy	1634	CTCCCTCCAAACTGGGTTCTTTGCCGGCTGTTTACGGACACAAGTTCAACTCGTCCGG	1693
Db	1625	CTCTCTCCAGACTGGGTTTCTTGGCCGGCTGTGTTACACACATAGTTTCAACTTCGTCGG	1684
Qy	1694	GTGCCCGGAGCGCATGGCCAGCTGCCGCCCATTTGACTGGTTCGCCCCAGGGGTGGGGCCC	1753
Db	1685	GTGCCCAAGCGCATGCCAGTCCAGTCCCAATGACAAAGTTCCACAGGATGGGGTCC	1744
Qy	1754	CATCACCTTACTAAGCCTTAACAGCTCGGAATCAGAGCCCTTATGTCTGGCAATTACGGGCC	1813
Db	1745	CATTACTTATGCTGAGTCTAGCAGATCAGACCCAGAGCCATATTGCTGGCACTTACCACACC	1804
Qy	1814	TCGACCGTGTGGTGTCTGCTACCGCGTCGCAAGTGTGTGGTCCAGTGTATTTGTTTCCACCCC	1873
Db	1805	TCCACAAATGACAAATCGTACTGCGTTCGGAGGTGTGGGCCCAAGTACTGCTTCCACCCC	1866
Qy	1874	AAGCCCTGTTGTGGTGGGGACCAACCATTCGTTCCGGGTGCCCTACGTATAGCTGGGGGGA	1933
Db	1865	AAGCCCTGTCTGTGGGACGACCGGATCGTTCCGTGTCCCTACGTATAGCTGGGGGGA	1924
Qy	1934	GAATGAGACAGACGTGATGCTCTCTAAACAACGCGTCCGCCACAAGGCACACTGGTTCGG	1993
Db	1925	GAACGAGACTGACGTGTCTGTCTCAACAACAACGCGCGCCCGCAAGCAACTGGTTCGG	1984
Qy	1994	CTGTACATGATGAATAGTACTGGGTTCACTAAGACGTGCGGAGGTCGCCCGTGTAAACAT	2053
Db	1985	CTGGCAATGATGAATAGCACCGGGTTCCACAACAATGTGGGGGGCCCCGGTGTAAACAT	2044
Qy	2054	CGGGGGGTGGTAAACCGCACTTGATCTGCCCCACGAGCTGCTCCGGAAGCACCCCGGA	2113
Db	2045	CGGGGGGTGGGCAACAACACCTGACCTTGCCCCACGGACTGCTTCGGAAGCACCCCGGA	2104
Qy	2114	GGCTACTTACAAAATGTGGCTCGGGGCCCTGGATTGACACTTGTAGTGTGCTAGTGAACATA	2173
Db	2105	GGCTTACCTACAAAATGTGGTTTGGGGCCCTTGGCTGACACTTGTAGTGTGCTAGTGAACATA	2164
Qy	2174	CCCATACAGGCTTGGCACTTACCCCTGCACCTCAATTTTCCATCTTTTANGGTTAGGAT	2233
Db	2165	TCCATACAGGCTCTGGGATTTACCCCTGCACCTGTAACTTTTACCATCTTCAAGGTTTAGGAT	2224
Qy	2234	GTATGTGGGGGGGTGGAGCACAGGCTCAATGGCGATGCAATTTGGCACTCGAGGAGAGCG	2293
Db	2225	GTATGTGGGGGGGTGGAGCACAGGCTCAATGTGCAATGCAATTTGACCCGAGGAGAGCG	2284
Qy	2294	CTGTAACTTGGAGGACAGGATAGTTCAGAACTCAGCCCGCTGCTGCTGTCTACACACAGA	2353
Db	2285	TTGTGACTTGGAGGACAGGATAGCCGGAGCTCAGCCCGCTGCTGCTGTCTACACACAGA	2344
Qy	2354	GTGSCATACTGCCCTGTCTTTTACACCCCTACCGGCTTTATCCACTGGTTTGTATCCA	2413
Db	2345	GTGGCAGTACTGCCCTGTCTTTTACACCCCTACCGCTTGTCTCAGTGGCTTGTATCCA	2404
Qy	2414	TCTCCATCAGAAACATCGTGGACCTGCAATACCTGTACGGTGTAGGCTCAGCGTTGTCTTC	2473
Db	2405	CCTCCATCAGAAACATCGTGGACCTGCAATACCTGTACGGTGTAGGCTCAGCGTTGTCTTC	2464
Qy	2474	CTTTGCAATCAAAATGGAGTACATCCCTGTGTCTTTCTCTCGGACAGCGCGGT	2533
Db	2465	CTTTGCAATCAAAATGGAGTATGCTCTGTGTCTTTCTCTCTCTAGCGGACGCACTGT	2524
Qy	2534	GTGTGCTGCTTGTGATGATGCTGTCTGTATAGCCACAGGCTGAGGCGGCTTACAGAACTT	2593
Db	2525	CTGTGCTGCTTGTGATGATGCTGTCTGTATAGCCACAGGCGGAGCGGCTTGGAGAACTT	2584
Qy	2594	GGTGGTCTCAATGGCGGCTCGGTGGCGGAGGATGATTTCTCTCTCTTCTTCTGTGT	2653
Db	2585	GGTGGTCTCAATGGCGGCTCTGTGGCCGGGACATGGCATCTCTCTCTCTCTTGTGT	2644
Qy	2654	CTTCTGCGCGCCTGTGATTAAGGGCAGGCTGGCTCTGGGGGGCGGTATGCTTTTATA	2713

D5	7025	CCTGTGGCGG	CAGAGATGGCGGGAACT	CACCCGGTGGAGTCGGAGAACAGGTGGT	7084
QY	7094	AATTTCTGAC	TCTTTTCGAACCGCTT	CACCGGAGGGGATCAGAGGGAGATATCCGTCCG	7153
D5	7085	AGTCTGTGACT	CTTTTGGACCCGCTT	CGAGCGAGGAGATCAGAGGGATATTCCTGTTCC	7144
QY	7154	GGCGGAGAT	CTCTGCAANAATCCAGGAGT	TTCCTCTAGGGTTGCCATATGGGCACGCCC	7213
D5	7145	GGCGGAGAT	CTCTGGGAAATCCAGAAAT	TTCCTCGCAGATGCCATCTGGGCGCGCCC	7204
QY	7214	GGACTACAA	TCTCCTACCTGTAGAGT	CTCTGAGGACCGGACTACCTCCCTCGGTGGT	7273
D5	7205	GGATACAA	CCCTTCACTGTTAGAGT	CTCTGGAAGAACCGGACTACCTCCCTCGGTGGT	7264
QY	7274	ACAGGATGCC	AAATGGCCACTTACCAAGGCT	TCTCCAATACCACTCCACGGAGAAAGAG	7333
D5	7265	GCACGGTGT	CGCGTGCACCTATCAAGGCC	CTCCATACCACTCCACGGAGAAAGAG	7324
QY	7334	GACGGTGT	CTCTGACAGAAATCCAAAT	TGTCTTCTGGCTTGGCGGAGTCCCACTAAGAC	7393
D5	7325	GACGGTGT	CTCTAAACAGAGTCTCCGTT	GTCTTCTGCTTAGCGGAGTCTACTAAGAC	7384
QY	7394	CTTCGGTAG	CTCCGATCTCTCGGCGCT	TGATAGCGCACCGGACCGCCCTTCTGACCT	7453
D5	7385	CTTCGGCAG	CTCCGAATCATCGGCCGTC	GCAGCGGACCGCGACCGCCCTTCTGACCA	7444
QY	7454	GGCTCCGAG	CAAGGTCAAAAGATCCGAGT	TGAGTCTTCTCTCCATCCGCCCCCT	7513
D5	7445	GGCTCCGAG	CAAGGTCAAAAGATCCGAGT	TGAGTCTTCTCTCCATCCGCCCCCT	7504
QY	7514	TGAAGGGAG	CGCGGACCCCGAATCTAC	GGAGCGGTCTTGGTCTACCGTGTGAGGA	7573
D5	7505	TGAGGGGAA	ACCGGGGACCCGATCTCAG	TGACGGGTCTTGGTCTACCGTGTGAGGA	7564
QY	7574	GGCTAGTAG	AGAGTCTCTCTGCTCTCA	ATGTCTTATACGTGTGAGGAACTCTTTGCTCG	7633
D5	7565	AGCTAGTAG	AGAGTCTCTCTGCTCTCA	ATGTCTTATACGTGTGAGGAACTCTTTGCTCG	7624
QY	7634	GCCATCGCT	TCGGAGGAAAGTAAAGT	TGCCCATCAACCCGCTTACAGGAACTCTTTGCTCG	7693
D5	7625	GCCATCGCT	TCGGAGGAAAGTAAAGT	TGCCCATCAACCCGCTTACAGGAACTCTTTGCTCG	7684
QY	7694	TCACCAACA	ATGGTCTACGCCACAACA	TCCCGAGCGCAAGCTCTCGGAGAGAGGT	7753
D5	7685	CCACCAAT	ACATGGTTTATGCCACA	ATCTCGCAGCGGAGCTCTCGGAGAGAGGT	7744
QY	7754	CACCTTTG	ACAGATTCGACGCTCTG	ATGATCATACCGGAGCTCTCAGGAGATGAA	7813
D5	7745	CACCTTTG	ACAGATTCGACGCTCTG	ATGATCATACCGGAGCTCTCAGGAGATGAA	7804
QY	7814	GGCGAAGG	CGTCCACAGTTAAGCT	TTCATCTATAGAGAGGCTCTGACGCTGAC	7873
D5	7805	GGCGAAGG	CGTCCACAGTTAAGCT	TTCATCTATAGAGAGGCTCTGACGCTGAC	7864
QY	7874	GGCCCCACA	TTTCGGCAAATCCAAAT	TTCGCTATGGGGCAAGGAGCTCGGAACCTATC	7933
D5	7865	GGCCCCACA	TTTCGGCAAATCCAAAT	TTCGCTATGGGGCAAGGAGCTCGGAACCTATC	7924
QY	7934	CAGCAGGG	CGGTTAACACATTCGCT	GTGGAGGACTTCTCTGGAAGACACTGAAC	7993
D5	7925	CAGCAGGG	CGGTTAACACATTCGCT	GTGGAGGACTTCTCTGGAAGACACTGAAC	7984
QY	7994	ACCAATTGA	CACCACCACCATCATG	GCACAAAGTGTCTGCTCTCAACCGAGAGAGG	8053
D5	7985	ACCAATTGA	CACCACCACCATCATG	GCACAAAGTGTCTGCTCTCAACCGAGAGAGG	8044
QY	8054	AGSCCGA	AGCCAGCTCGCTTATCG	TATTCACAGACCTGGGAGTTCGTGTATGCGAGAA	8113
D5	8045	AGSCCGT	TAAGCCAGCCGCTTAT	CGTATTCACAGATCTGGGAGTCCGTGTATGCGAGAA	8104
QY	8114	GATGGCC	CTTTACGAGTGGTCTCC	ACCTTCTCTAGGCCGTGATGGGTCTCTCATAGG	8173

Db	8105	GAIGGCCCTCTATGATGTGGTCTCCACCCCTTCTCTCAGTCTGTGTAGTGGCTCCTCATACGG	8164
QY	8174	ATTTCATACTCCCCAAGCAGCGGGTCGAGTTCCTCGTGAATAACTCTGGAAATCAAGAA	8233
Db	8165	ATTCCAGTACTCTCTGGGCAGCGAGTCGAGTTCCTGGTGAATACCTTGGAAATCAAGAA	8224
QY	8234	ATGCCCTATGGCTTCTCATATGACACCCCGCTGTTTTGACTCAACGGTCACTGAGAGTGA	8293
Db	8225	AAACCCCATGGCTTTTCATATGACACTCGCTGTTCGACTCAACGGTCAACCGAAGCA	8284
QY	8294	CATTCTGTGTTGAGAGTCAATTTACAAATGTTGTGACTTGGCCCGCCGAGCCACAGAGC	8353
Db	8285	CATCCGTGTTGAGAGTCAATTTACCAATGTTGTGACTTGGCCCGCCGAAACCCACAGAGC	8344
QY	8354	CATAAGTTCGCTCACAGAGCGGTTTACATCGCGGGGTCCCTTCACACTAACTCAAAAAGGCA	8413
Db	8345	CATAAATCGCTCACAGAGCGCTTTATATCGGGGGTTCCTCTGACTAAATTCAAAAGGGCA	8404
QY	8414	GAACTGGGTTATCGCCCGTCCGCGCAAGTCGCTGACGACTAGTTCGGTGAATAC	8473
Db	8405	GAACTGGGTTATCGCCCGTCCGCGAGCGCGTCTGACGACTAGTTCGGTGAATAC	8464
QY	8474	CCTCACATGTTACTTTGAAGCGCACTCAGAGCTGTGAGCTGTGAAAGTCCAGACTTGCAC	8533
Db	8465	CCTCACATGTTACTTTGAAGCGCTCTGAGCGCTGTGAGCTGTGCAAGCTCAGGACTTGCAC	8524
QY	8534	GATGCTCGTGAAACGAGACGACCTTGTGCTGTTACTGTGAAAGCGGGACCCAGAGGA	8593
Db	8525	GATGCTGTGAAACGAGACGACCTTGTGCTGTTACTGTGAAAGCGGGACCCAGAGGA	8584
QY	8594	TGCGCGCGCCTACGAGCCTTTCAGGAGGTATGACTAGTATTCGCGCCCGCCCGGGGA	8653
Db	8585	CGCGCGAGCCTACGAGTCTTTCAGGAGGCTATGACTAGTATTCGCGCCCGCCCGGGGA	8644
QY	8654	TCGCGCCCAACAGATACGACTTGGAGCTGTAAATCATGTTCTCCAAATGTGTCAGT	8713
Db	8645	CCGCGCCCAACAGATACGACTTGGAGCTGTAAATCATGTTCTCCAAATGTGTCAGT	8704
QY	8714	CGCGCAGATGCTCTGGCAAAAAGGTATACCTCAACCGTGACCCACACACCCCTCT	8773
Db	8705	CGCGCCACATGCTACGGAANAAGGTGTACTACCTCACCGGTGATCCACACACCCCTCT	8764
QY	8774	TGCAACGGGCTGCGTGGGAGACAGCTAGACACACTCCAAATCAACTCTTGGCTAGGCAAT	8833
Db	8765	AGCACGGGCTGGTGGGAGACACTAGACACACTCCAGTACTCTTGGCTAGGCAAT	8824
QY	8834	CATCATGTATGGCGCACCCCTATGGCGAAGATGATTCGTGATGACTCACTTTTCTCCAT	8893
Db	8825	TATATGTATGGCGCCACTTTTGGGCAAGGATGATTCGTGATGACTCACTTTTCTCCAT	8884
QY	8894	CTTTCTAGCTCAAGAGCACTTGAABAGCCCTGGATTCGTAGATCTAGGGGCTTGCTA	8953
Db	8885	CTTTCTAGCGCAGAGCACTTGAABAGCCCTGGACTCCAGATCTAGGGGCTTGCTA	8944
QY	8954	CTCAATTGACCACTTGACCTACCTCAGATCATTTGAACGACTCCATGGTCTTAGCGCAT	9013
Db	8945	CTCAATTGACCACTTGACCTACCTCAGATCATTTGAACGACTCCATGGTCTTAGCGCAT	9004
QY	9014	TACACTCCACAGTTACTCTCCAGGTGAGATCATATAGGTGGCTTCATGGCTCAGAGAACT	9073
Db	9005	TTCACTCCATAGTTACTCTCCAGGTGAGATCAATAGGTGGCTTCATGGCTCAGAGAACT	9064
QY	9074	TGGGTACCACTCTGGCACTGGAGACATCGGCGCAAGTGTCCCGCTTAAGCTACT	9133
Db	9065	TGGGTACCACTCTGGAGTCTGGAGACATCGGCGCAAGTGTCCCGCTTAGGCTACT	9124
QY	9134	GTCCCGAGGGGGAGGGCGCCCACTTGTGGCAGATACCTCTTTAACTGGGAGTAAAGAC	9193
Db	9125	GTCCCGAGGGGGAGGGCGCCCACTTGTGGCAGATACCTCTTTAACTGGGAGTAAAGAC	9184
QY	9194	CAGACTTAACTCATCTCAATCCCGCCCGCTCCAGCTTGTCTGTGCTGGTTCGT	9253
Db	9185	CAAACTTAACTCATCTCAATCCCGCTCGCTCCCGGTGGAATGTCCGGGGTTCGT	9244

QY 1081 CGGCGAGGAATGCCACGCTCCCACTACGACAAATACGACGCCACGCTGCTGCTG 1140
DB 1082 CGGCCAGGAATGCCACGCTCCCACTACGACAAATGCGACGCCATGCTGCTGCTG 1141
QY 1141 GGAAGGCTCTCTTCTGCTCGCTATGATAGTGGGGATCTCTGCGGATCTATTTCTCTG 1200
DB 1142 GGGTAGCTCTCTTCTGCTCGCTATGATAGTGGGGACCTCTGCGGATCTGTTTCTCTG 1201
QY 1201 TCTCCAGCTGTTCACCTTCTCGCTCGCGGATGAGACAGTGCAGGACTGCAACTGCT 1260
DB 1202 TTTCCAGCTGTTCACCTTCTCGCTCGCGGATGAGACAGTGCAGGACTGCAACTGCT 1261
QY 1261 CAATCTATCCCGGCAATGATCAGGTCAACGCAATGCTTGGGATGATGATCAACTGCT 1320
DB 1262 CAATCTATCCCGGCGGATGATCAGGTCAACGCAATGCTTGGGATGATGATCAACTGCT 1321
QY 1321 CACCTAACACAGCCCTAGTGGTGTGCGAGTCTCGGATCCCAAGTGTGCTGGACA 1380
DB 1322 CGCCTAACACAGCCCTAGTGGTGTGCGAGTCTCGGATCCCAAGTGTGCTGGACA 1381
QY 1381 TGGTGGCGGGGCCCACTGGGAGTCTTGGCGGGCTTGGCTACTATTCCATGTTAGGGA 1440
DB 1382 TGGTGCACAGGTCCTCACTGGGGAATCCCTGGCGGGCTTGGCTACTATTCCATGTTAGGGA 1441
QY 1441 ACTGGGCTAAGGTTCTGATTGGGCTACTCTTTGGCGGCTTGAAGGAGACCCACA 1500
DB 1442 ACTGGGCTAAGGTTCTGATTGGGCTACTCTTTGGCGGCTTGAAGGAGACCCACA 1501
QY 1501 CGACGGGAGGTTGGCGGCGCACACACCTCGGGTTCAGGTCCTTTCTCATCTGGGG 1560
DB 1502 TGACAGGGGGGCGCAAGTGGGCGCTAGTCTGCTAACGTCCTTTTACGCTGGGC 1561
QY 1561 CGTCTCAGAAATCCAGCTTGTGAATACCAACGGCAGTGTGCATCAACAGGACTGCC 1620
DB 1562 CGGTTACAGCCTCCAGCTCTAACAACCAACGGCAGTGTGCATCAACAGGACCGCC 1621
QY 1621 TAATTTGAATGACTCCCTCAAACTGGGTTCTTTGGCGGCTTGTTCAGACACAAGT 1680
DB 1622 TGAGTGAATGACTCCCTCAAACTGGGTTCTTTGGCGGCTTGTTCAGACACAAGT 1681
QY 1681 TCAACTCTCGGGTGGCGGAGCGCATGGCAGCTGGCGGCGCCCATCTGCTGCTGCTG 1740
DB 1682 TCAACGCTTCGGGTGGCGGAGCGCTTGGCACGTCGCGGCGCCCATGATCAATTCGGC 1741
QY 1741 AGGGTGGGGCCCATCTAATTAAGCTTAACAGCTCGGATCAGAGGCTTATTTGCT 1800
DB 1742 AGGGTGGGGTCCCATCACTTACACTGAGCCTCATGATTGGATCAGAGGCTTATTTGCT 1801
QY 1801 GGCATTACGCGCTCGACCGTGGTGTGCTACCGGTCGACAGTGTGCTCAGTGT 1860
DB 1802 GGCATTACGCGCTCAACCGTGGTGTGCTACCGGTCGACAGTGTGCTCAGTGT 1861
QY 1861 ATTTTTCACCCCAAGCCTCTTGTGGTGGGACCAACGATCTGTCGCTGCTGCTGCTGCT 1920
DB 1862 ACTGCTTACCCCGAGTCTGTTGGGTGGGACTACCGATCTGTTGGTCCCTCAT 1921
QY 1921 ATAGTGGGGAGATGAGACAGCTGATGCTCTCAACCAACGCTCGGACCAAG 1980
DB 1922 ACAGATGGGGGCAATGAGACGACGCTGCTGCTCTTAAACAGCGCGGCGCGCAAG 1981
QY 1981 GCAACTGCTTGGCTGTACATGAGTGTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
DB 1982 GCAACTGCTTGGCTGTACATGAGTGTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCT 2041
QY 2041 CCCGCTGTACATCGGGGGTGTGCTAACCGCACTTGTGCTGCTGCTGCTGCTGCTGCTGCT 2100
DB 2042 CCCGCTGTACATCGGGGGTGTGCTAACCGCACTTGTGCTGCTGCTGCTGCTGCTGCTGCT 2101
QY 2101 GGAAGCACCCGAGGCTACTTACAAATGCTGCTGGGCTGCTGGTGAACCTAGGT 2160
DB 2102 GAAAGCACCCGAGGCTACTTACAAATGCTGCTGGGCTGCTGGTGAACCTAGGT 2161
QY 2161 GCCTAGTAGACTACCCATACAGGCTTTGGCACTACCCCTGCATCTCAATTTTCCATCT 2220

DB 2162 GCTTASTCGACTACCGGTACAGGCTCTGCGATTACCCCTGCACTGTCAACTTACCATCT 2221
QY 2221 TTAAGTTAGATATGATGCGGGGCGTGGAGCACAGGCTCAATCCCGCATGCAATTTGA 2280
DB 2222 TTAAGTTAGATATGATGCGGGGCGCGGAGCACAGGCTCGACCGCATGCAACTTTGA 2281
QY 2281 CTGAGGAGAGCGCTGAATCTGAGGACAGGATAGGTAGAGCTCAGCCGCTGCTGCT 2340
DB 2282 CTGAGGAGAGCGCTGATGACCTGAGGACAGGATAGGTAGAGCTCAGCCGCTGCTGCT 2341
QY 2341 TGTCTAACACAGAGTGGGAGATGACCTGCGCTGCTGCTTCCACCACTTACCGCTTATCCA 2400
DB 2342 TGTCTAACACAGAGTGGGAGATGACCTGCGCTGCTGCTTCCACCACTTACCGCTTATCCA 2401
QY 2401 CTGTTTGAATCACTCTCCATCAGAACATCTGTTGAGTGAATCACTCTGATAGCCAGGCTCAGGCG 2460
DB 2402 CTGTTTGAATCACTCTCCATCAGAACATCTGTTGAGTGAATCACTCTGATAGCCAGGCTCAGGCG 2461
QY 2461 CAGCGTTTGTCTCTTTCGAAATCAATGGGAGTACATCCCTGCTGCTTCTCTCTCTCTG 2520
DB 2462 CGGCGTTTGTCTCTTTCGAAATCAATGGGAGTATATTTGCTGCTCTCTCTCTCTCTCTG 2521
QY 2521 CAGACGCGCGGCTGTGCTGCTTGTGAGTGAATGCTGCTGATAGCCAGGCTCAGGCG 2580
DB 2522 CGGACGCGCGGCTGTGCTGCTTGTGAGTGAATGCTGCTGATAGCCAGGCTCAGGCG 2581
QY 2581 CTTTAGAGAACTTGGTGTCTCTCAATCGGGGCTGCTGCGGAGCGCATGATGCTCT 2640
DB 2582 CTTTAGAGAACTTGGTGTCTCTCAATCAGGCTGCTGCGGAGCGCATGATGCTCT 2641
QY 2641 CTTTCTTGTCTCTCTGCGCGCTGCTACATTAAGGAGGCTGCTGCTCTCTCTCTCTCTCTG 2700
DB 2642 CTTTCTTGTCTCTCTGCGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2701
QY 2701 CGTATGCTTTTATGGGCTATGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
DB 2702 CATACGCGCTCTATGGCGTATGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2761
QY 2761 CTTAGCGCTTGGACCGGAGATGCTGCTGCTGCGGGTCTGCTGCTCTCTCTCTCTCTCTCT 2820
DB 2762 CGTACGCGCTTGGACCGGAGATGCTGCTGCTGCGGGTCTGCTGCTCTCTCTCTCTCTCTCT 2821
QY 2821 TATTTCTTGAATCTGCTACCATCTACAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
DB 2822 TACTTTTGAATCTGCTACCATCTACAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2881
QY 2881 AATCTTTTATACAGAGCGGAGCGCATCTGCAAGTGTGGGTCTCTCTCTCTCTCTCTCTCTCT 2940
DB 2882 AATCTTTTATACAGAGCGGAGCGCATCTGCAAGTGTGGGTCTCTCTCTCTCTCTCTCTCTCT 2941
QY 2941 GGGAGCGCGGATGCTCATCT 3000
DB 2942 GGGAGCGCGGATGCTCATCT 3001
QY 3001 ACATCAACAACT 3060
DB 3002 ACATCAACAACT 3061
QY 3061 CGAGAGTGGCTTCT 3120
DB 3062 CTAGAGTGGCTTCT 3121
QY 3121 AAGTTCGCGGGGCT 3180
DB 3122 AAGTTCGCGGGGCT 3181
QY 3181 CGTACGCTTTTATACCACT 3240
DB 3182 CGTACGCTTTTATACCACT 3241
QY 3241 TTGCGGTGGGCTAGAGCGGCT 3300

Db 3242 TTGCAATGSCAGTACAGACCGCTGTCTTCTCTGACATGAGACCAAAATCATCACTGGG 3301
 Qy 3301 GAGCAGACACCGCTGCTGGGACATCATCTTGGGTCTACCGTCTCCGCCGAAAGG 3360
 Db 3302 GGGCAGACACCGCGGTGCGGGACATCATCTTGGCTGCGCTCTCCGCCGAAAGG 3361
 Qy 3361 GGAAGGAGATATTTTGGACCGGTGTAGTCTCGAAGGGCAAGGGTGGGACATCCCTTG 3420
 Db 3362 GGAAGGAGATATTTTGGACCGCGGATAGTCTTGAAGGACAGGGTGGGACATCCCTTG 3421
 Qy 3421 GCGCCATACCGGCTACTTCCCAAAACCGCGGGGTACTTGGTTCATCATCACTAGCC 3480
 Db 3422 GCGCCATACCGGCTACTTCCCAAAACCGCGGGGTACTTGGTTCATCATCACTAGCC 3481
 Qy 3481 TCACAGCGCGGACAAAGAACAGGTTCGAAGGGAGGTTCAGTGGTTCATCACTAGCC 3540
 Db 3482 TTACAGCGCGGACAAAGAACAGGTTCGAAGGGAGGTTCAGTGGTTCATCACTAGCC 3541
 Qy 3541 AATCTTTCTGCGACCTGCTCATCAACGGCGTGTGCTGACTGTCTACCATGGCGTGGCT 3600
 Db 3542 AATCTTTCTGCGACCTGCTCATCAACGGCGTGTGCTGACTGTCTTCCAGCGCGCGCT 3601
 Qy 3601 CGAAGACCTAGCGGTTCGAAAGGTCCAAATCACCCAAATGTACACCAATGTAGACCTTG 3660
 Db 3602 CAAGACCTTAGCGCGCTTCAAGGGTCCAAATCACCCAAATGTACACCAATGTAGACCAAG 3661
 Qy 3661 ACCTCGTGGGTGGCAGCGCCCGCGGGCGCTCCATGACACCATGACAGCTGTGCA 3720
 Db 3662 ACCTGTGGGTGGCGCGACCTCTGGGGCGGTTCCTGACACCATGCACTTGGCGCT 3721
 Qy 3721 GCTCGGACTTACTTGGTTCAGACATGTGTATGTCATTCGGGTGCGCGCGCGGCG 3780
 Db 3722 CCTCGGACTTACTTGGTTCAGACATGTGTATGTCATTCGGGTGCGCGCGCGGCG 3781
 Qy 3781 ACAGCAGGGAGTCTACTCTCCCAAGCGCGTCTCCTACTGAAAGGTCTCTGGGTG 3840
 Db 3782 ACGGTAGGGGAGCTACTCTCCCGCGCGCTCTCTACTTGAAGGGTCTCTCGGGTG 3841
 Qy 3841 GTCCATTTCTTGGCTTGGGGACGTCGTGGGGTCTTCCGGTCTCTGAGGGTCTCTCGGGTG 3900
 Db 3842 GTCCACTGCTCTGCGTTCGGGGACGCTGCGGATACTTCCGGTCTCTATGCAACC 3901
 Qy 3901 GGGGGTGGGAAGCGGTGACTTCATPACCGTGTAGTCTATGAAACTACCATGCGGT 3960
 Db 3902 GGGGGTGGGAAGCGGTGAAATCAATACCGTGTAGTCTATGAAACTACCATGCGGT 3961
 Qy 3961 CTCGGTCTTACAGACACTCAACCCCGCGGTGTACCGACATTCOAATGAGGAC 4020
 Db 3962 CTCGGTCTTACAGACACTCAACCCCGCGGTGTACCGACATTCOAATGAGGAC 4021
 Qy 4021 ATCTGCAAGCTCTTACTGGCAGCGGCAAGACCCMAAGTCCCGTCTGCTATGAGGCC 4080
 Db 4022 ACTTACAGCTCTCCACCGGACGCGGACAGCATAGGGTCCCGGTCTCATATGAGCCCC 4081
 Qy 4081 AAGGTACAGGTGCTGCTCTGAAACCGGTCTGCGCGCACCTTAGGGTTTGGGGCGT 4140
 Db 4082 AAGGTACAGGTGCTGCTCTGAAATCGGTCTGCGCGCACCTTAGGGTTTGGGGCGT 4141
 Qy 4141 ATATGTCCAAGGCACAGGTATCCACCTTACATCAGACTGGGGTAAGACCATTACCA 4200
 Db 4142 ATATGTCCAAGGCACATGGTATCCACCCCACTTAGAATCGGGTAAGGACCATACCA 4201
 Qy 4201 CGGCGGCTTCAATACGTACTTCCACTATGGCAAGTTCTTGGCGACGGTGGCGTCTG 4260
 Db 4202 CAGGTGCCCTTATCACAATCTCCACTATGCAAGTTCTTGGCGACGGTGGCGGCTCCG 4261
 Qy 4261 GGGGGGCTTATGATCATATATATGTGATAGTGCCTCACTCACTGACTCGACTACCATCT 4320
 Db 4262 GGGGGGCTTATGATCATATATATGTGATAGTGCCTCACTCACTGACTCGACTACCATCT 4321
 Qy 4321 TGGGCATCGGCACAGTCTTGGCAAGCGGACAGGCTTGGAGCGCGGTCTGCTGCTG 4380
 Db 4322 ATGGCATCGGCACAGTCTTGGCAAGCGGACAGGCTTGGAGCGCGGTCTGCTGCTGCT 4381

Qy 4381 CCAACCGCTACACCTCCGGATTCGGTTCACCGTGCACACCCCAATATCGAGGAATAGGCC 4440
 Db 4382 CCAACCGCTACACCTCCGGATTCGGTTCACCGTGCACACCCCAATATCGAGGATGGGCC 4441
 Qy 4441 TGTCCAAACATGAGAGATCCCTTCTATGGCAAGCCATCCCATTCAGGCCATCAAGG 4500
 Db 4442 TGTCTAATATCTGGAGATCCCTTCTACGCAAGCCATTCCTCATCGAGGTATCAAGG 4501
 Qy 4501 GGGGAGGACATCTCATTTCTTGGCCATTCCTCAAGAAAGAAATGACGAGCTCGCCGCAAGC 4560
 Db 4502 GGGGAGGACATCTCATTTCTTGGCCATTCCTCAAGAAAGAAATGACGAGCTCGCCGCAAGC 4561
 Qy 4561 TGACAGGCTCGGACTGAAAGCTGTACATATATACCGGGCCCTTGAATGTCGCTCATAC 4620
 Db 4562 TGTGAGGCTCGGACTCAATGCGCTAGCTATTAACCGGGTCTTGAAGTCTCGCTCATAC 4621
 Qy 4621 CGCCTATCGGAGACGCTGCTGTCGCAACAGACGCTCTAATCAAGGGTTTCAACCGCG 4680
 Db 4622 CGACAGGAGAGCTGTTGTCGTCGACGACGCTCTAATGACGGGTCTTACCGGCG 4681
 Qy 4681 ATTTGACTCAGTGAATGCAATACATGTGTCAACCGACAGCTCCTTCACTTGG 4740
 Db 4682 ACTTTGACTCAGTGAATGCAATACATGTGTCAACCGACAGCTCCTTCACTTGG 4741
 Qy 4741 ATCCACCTTCAACATTCAGACGACGCTGTCGCGCCGACGCGGTGTCGCGTTCGCAAC 4800
 Db 4742 ACCCACCTTCAACATTCAGACGACGCTGTCGCGCCGACGCGGTGTCGCGTTCGCAAG 4801
 Qy 4801 GGCAGGTGAACCTGGCAGGGTAGGAGTGGCATCTACAGGTTTGTGACCTCCAGGAGAAC 4860
 Db 4802 GGCAGGACGACCTGGTAGGGGACGGCTGGCATATACAGTTTGTGACTCCAGGAGAAC 4861
 Qy 4861 GGCCTCGGCGATGTCGATTTCTTGGTCTGTGTGAGTGTATGACGCGGGTGTGCTT 4920
 Db 4862 GGCCTCGGCGATGTCGATTTCTTGGTCTGTGTGAGTGTATGACGCGGGTGTGCTT 4921
 Qy 4921 GGTATGAGTCAACCGCGTGTGAGCTCGGTGTAGTTGCGGGTTACCTAAAATACACAG 4980
 Db 4922 GGTACGAACTCAGCGCGCTGAGACCTCGGTGTAGTTGCGGGTGTGCTTAAACACACAG 4981
 Qy 4981 GGTTCGCGCTGTGCGACGACCATCTGAGTTCCTGGAGAGCTCTTACAGAGGCTCACCC 5040
 Db 4982 GGTTCGCGCTGTGCGACGACCATCTGAGTTCCTGGAGTTCCTGAGGAGTCTTCAAGGCTCACCC 5041
 Qy 5041 ACATAGTCCCACTTCTCTGCGACGACTAAAACAGGACGAGAGCAACTTTCTTACCTGG 5100
 Db 5042 ACATAGTCCCACTTCTTATCCAGACTAAAACAGGACGAGAGCAACTTTCTTACCTGG 5101
 Qy 5101 TGCCATATCAAGCTACAGTGTGCGCAGGCTCAAGCTCCACCTGATGTCGAGCAAAA 5160
 Db 5102 TAGCATACAGGCTACAGTGTGCGCAGGCTCAAGCTCCACCTGATGTCGAGGATGAA 5161
 Qy 5161 TGTGGAAGTGTCTATACAGGCTGAAACCTACATGCACTGCAAGGGCCCAACACCTCTCTATA 5220
 Db 5162 TGTGGAAGTGTCTATACAGGCTGAAACCTACATGCACTGCAAGGGCCCAACACCTCTCTATA 5221
 Qy 5221 GGTAGAGCCGTCCAAATAGAGTCACTCTCAACACCCCAATAAATATCATCATGG 5280
 Db 5222 GGTAGAGCCGTCCAAATAGAGTCACTCTCAACACCCCAATAAATATCATCATGG 5281
 Qy 5281 CATGATGCGGTGACCTGGAGTGTCTACTAGCACTGCGGTGTGTTAGCGGAGTCC 5340
 Db 5282 CATGATGCGGTGACCTGGAGTGTCTACTAGCACTGCGGTGTGTTAGCGGAGTCC 5341
 Qy 5341 TTGCAAGTGTGCGCGATATCTGCTCAGCAGGAGTGTGTTGTTGTTGGGAGATCA 5400
 Db 5342 TCGAGCTCTGGCGGTACTGCTGCAACAGGACGCGGTGTTGTTGTTGGGAGATCA 5401
 Qy 5401 TCTTGTTCGGGAAGCAGCTGTCTTCCGACAGGGAAGTCTCTTACAGGAGTTCGATC 5460
 Db 5402 TCTTGTTCGGGAAGCAGCTATCATCCCGATAGGGAAGTCTCTTACAGGAGTTCGACG 5461

QY 5461 AGATGGAAGAGTGTGCTTACAACTTCCCTTATCATCGAGCAGGAAATGACGCTCGCCGAGC 5520
DB |||||
QY 5462 AGATGAGAGTGTGCTTACAACTTCCCTTATCATCGAGCAGGAAATGACGCTCGCCGAGC 5521
DB |||||
QY 5521 AATTCAAGCAAAAGCGCTCGGGTGTGTCGAAACGGCCCAAGCAAGCGGAGGCTGCTG 5580
DB |||||
QY 5522 AATTCAAGCAAAAGCGCTCGGGTGTGTCGAAACGGCCCAAGCAAGCGGAGGCTGCTG 5581
DB |||||
QY 5581 CTCCTCGTGTGAGTCCCAAGTCCGAGCCCTTGCAGACCTTCTGGCGAGACACATGTGGA 5640
DB |||||
QY 5582 CTCCTCGTGTGAGTCCCAAGTCCGAGCCCTTGCAGACCTTCTGGCGAGACACATGTGGA 5641
DB |||||
QY 5641 AATTCAATGCGGAATACAGTACCTAGCAGGCTTATCCACTCTGCTGGAACCCCGCGA 5700
DB |||||
QY 5642 ACTTCAATAGTGGGATACAGTACTTGGCAGGCTTGTCCACTCTGCTGGAACCCCGCA 5701
DB |||||
QY 5701 TAGCATCATGATGGCATTTACAGCTTCTATCAGTACGCCGCTCACCAACCCAAACACCC 5760
DB |||||
QY 5702 TAGCATCACCGATGGCAATTCACAGCTCCATCACAGCCCGCTCACCAACCCAGCATACC 5761
DB |||||
QY 5761 TCCTGTGTTAACTCTTGGGGGATGGGTGGCTGCTGAGACCTTCTGGCGAGACACATGTGGA 5820
DB |||||
QY 5762 TCCTGTGTTAACTCTTGGGGGATGGGTGGCTGCTGAGACCTTCTGGCGAGACACATGTGGA 5821
DB |||||
QY 5821 CAGCTTTCGTGGGCGCGGCAATCGCCGGAGCGGCTGTTGGCAGCATAGGCTTGGGAAGG 5880
DB |||||
QY 5822 CAGCTTTCGTGGGCGCGGCAATCGCTGGAGCGGCTGTTGGCAGCATAGGCTTGGGAAGG 5881
DB |||||
QY 5881 TGCTCGTGAGCATCTTGGGGGCTATGGGCAAGGGGTAGCGCGGACACTGCTGGGCTTTA 5940
DB |||||
QY 5882 TGCTTGTGAGCATCTTGGGAGGTTATGGAGCAAGGGGTGCGGGGCGCACTTGTGGGCTTTA 5941
DB |||||
QY 5941 AGGTCAATGAGCGCGGAGTGCCTCCACCGAGCACTGTGTCAACTACTTCCCTGCCATCC 6000
DB |||||
QY 5942 AGATCATGAGCGCGAGATGCCCTTTCAGCGAGGACATGTGTCACTTCTCCCTGCCATCC 6001
DB |||||
QY 6001 TCCTTCCTGGTGCCTGCTGCTGCGGGTGTGTGCGCAGCAATACTGCGTGGCAGCTGG 6060
DB |||||
QY 6002 TTTCTCCCGTGCCTGCTGCTGCGGATGTGTGTCAGCAATATCTGCGTGGCATGTGG 6061
DB |||||
QY 6061 GCCCGGAGAGGGGCTGTGCAGTGTGATGAACCGGCTGATAGCGTTCGCTGCGGGGTA 6120
DB |||||
QY 6062 GCCCAGGGAAGGGGCTGTGCAGTGTGATGAACCGGCTGATAGCGTTCGCTGCGGGGTA 6121
DB |||||
QY 6121 ACCAGCTTCCCTTACGCACTATGTGCTGTGAGCGCAGCTGACAGCACTGTCAGTCTAGA 6180
DB |||||
QY 6122 ACCAGCTTCCCTTACGCACTATGTGCTGTGAGCGCAGCTGACAGCACTGTCAGTCTAGA 6181
DB |||||
QY 6181 TCCTCTAGCTTACCATCACTCACTGCTGAGCGCTCCACAGTGGATTATGAGG 6240
DB |||||
QY 6182 TCCTTTCAGCGCTCACCATCACTCAGCTGTGTGAAGAGACTCCACAGTGGATTATGAGG 6241
DB |||||
QY 6241 ACTGCTTACGCCATGCTCCGCTGCTGCTAGAGGATGTTGGGATTGGATATGACGG 6300
DB |||||
QY 6242 ACTGCTTACGCCATGCTTCCAGCTGCTGGTGAAGGAGATTGGGACTGGATCTGACCG 6301
DB |||||
QY 6301 TGTGACTGCTTCAAGACCTTGGCTTCCAGTCCAAACTCTTCCCGGGGTTACCGGAGTCC 6360
DB |||||
QY 6302 TGTGACTGCTTCAAGACCTTGGCTTCCAGTCCAAACTCTTCCCGGGGTTACCGGAGTCC 6361
DB |||||
QY 6361 CTTTCTCTGATGCAACCGCGGTTCAAGGAGTCTGCGGGGGGACGCGCATATGCAAA 6420
DB |||||
QY 6362 CTTTCTCTGATGCAACCGCGGTTCAAGGAGTCTGCGGGGGGACGCGCATATGCAAA 6421
DB |||||
QY 6421 CCACCTGCCATGCGGAGACAGATTCGCGGACATGTCAAAACGGTTTCCATGAGGATCG 6480
DB |||||
QY 6422 CCACCTGCCATGCGGAGACAGATTCACCGACATGTCAAAACGGTTTCCATGAGGATCG 6481
DB |||||
QY 6481 TAGGSCCTTAGAATCTGCGCAACACCTGCGCAGGAACGTTCCCATCAACGATACACCA 6540
DB |||||
QY 6482 TTGGSCCTTAAGAACCTGCGCAGACACAGTGTGTCGGGACATTTCCCATCAACGCGTACACCA 6541
DB |||||
QY 6541 CGGGAACCTTGACACCTTCCCGGGCGCCAACTATTCCAGGCGGCTATGGCGGGTGGCTG 6600
DB |||||

DB 6542 CGGGCCCTTGACACCTTCCCGGGCGCCAACTATTCCAAGGCAATGTGGAAGTGGCGC 6601
QY |||||
QY 6601 CTGAGAGTACGTGAGGTATTACGCGTGTGGGGGATTTCCACTACGTAGCGGCAATGACCA 6660
DB |||||
QY 6602 CTGAGAGTACGTGAGGTACGCGGTTGGGAGATTTTCACTACGTAGCGGCAATGACCA 6661
DB |||||
QY 6661 CTGACAAAGTAAAGTCCCATGCGAGGTTCCGGCCCCCGAAATTTCTTACAGGAGTGGATG 6720
DB |||||
QY 6662 CTGACAAAGTAAAGTCCCATGCGAGGTTCCGGCCCCCGAAATTTCTTACAGGAGTGGATG 6721
DB |||||
QY 6721 GAGTCCGCTTGCACAGGTACGCTCCGGCTGCACAACTCTTCTACGCGGAGACGTCAGT 6780
DB |||||
QY 6722 GAGTCCGCTTGCACAGGTACGCTCCGGCTGCAGACCTCTCTACGCGGAGGAGTCTAT 6781
DB |||||
QY 6781 TCCAGGTCCGCTCAACCAATATCTTGTGCGGTGCGAGCTCCCATGCGAGCCGAAACCGG 6840
DB |||||
QY 6782 TCCAGGTCCGCTCCACAGTACCTTGTGCGGTTCACAGCTCCCATGCGAGCCGAAACCGG 6841
DB |||||
QY 6841 AGCTAAAGTCTTACTTCCATGCTCACCGATCCCTCCCAATTAACAGACAGACGCTA 6900
DB |||||
QY 6842 ATGTAGCAGTCTCTCACTTCCATGCTCTGACCTCCCTCCCAATTAACAGACAGACGCTA 6901
DB |||||
QY 6901 AGCTTAGCTCTGCTAGAGGCTCTCCCGCTCTTTAGCCAGCTCATACGTAGCAGTCTGT 6960
DB |||||
QY 6902 AGCTTAGCTCTGCTAGAGGCTCTCCCGCTCTTTAGCCAGCTCTTACGTAGCAGTCTGT 6961
DB |||||
QY 6961 CTGCGCTTCTTTGAGGCGACATGCACTACCCACATGACTCCCGCGGACGCTGACCTCA 7020
DB |||||
QY 6962 CTGCGCTTCTTTGAGGCGACATGCACTACCCACATGACTCCCGCGGACGCTGACCTCA 7021
DB |||||
QY 7021 TCGAGGCCAACTCTTGTGGCGGACAGAGATGGCGGGAACATCACTCCGCTGAGTCTCAG 7080
DB |||||
QY 7022 TTGAGGCCAACTCTTGTGGCGGACAGAGATGGCGGGAACATCACTCCGCTGAGTCTCAG 7081
DB |||||
QY 7081 AGAATAAGTACTTAACTCTGAGCTCTTTCGAAACCGCTTTCACGCGGAGGGGATGAGAGG 7140
DB |||||
QY 7082 AGAATAAGTACTTAACTCTGAGCTCTTTCGAAACCGCTTTCACGCGGAGGGGATGAGAGG 7141
DB |||||
QY 7141 AGATATCGCTCGCGGAGATCTGCGAAATCAAGGAAGTTCCTTCCAGCTTGGCCCA 7200
DB |||||
QY 7142 AAATATCGCTTCCGCGGAGATCTGCGAAATCAAGGAAGTTCCTTCCAGCTTGGCCCA 7201
DB |||||
QY 7201 TATGCGGACGCGCGGACTTACAACTCTCCACTGTAGAGTCTGGAAGGACCGCGACTACG 7260
DB |||||
QY 7202 TATGCGGCGCGCGGATTAACCCCTCGCTGTGAGAGTCTTGAAGGACCGCGACTACG 7261
DB |||||
QY 7261 TCCCTCCGCTGTACAGGATGCGCATGCGCACCTTACCAAGGCTTCTTCCAAATACCACTC 7320
DB |||||
QY 7262 TCCCTCCGCTGTGTACAGGCTGCGGCTTCCGCGCCACCAAGGCGCTTCCAAATACCACTC 7321
DB |||||
QY 7321 CACGGAAGAGAGACGCTTGTCTGACAGATCCAAATGTGTCTTCTGCTTGGCGGAGC 7380
DB |||||
QY 7322 CACGGAAGAGAGACGCTTGTCTGACAGATCCAAATGTGTCTTCTGCTTGGCGGAGC 7381
DB |||||
QY 7381 TCGCACTTACAGCTTCCGCTGCTGCGGCTGTGATAGCGGACCGCGGAGC 7440
DB |||||
QY 7382 TCGCTATAGACCTTCCGCTGCTGCGGCTGTGATAGCGGACCGCGGAGC 7441
DB |||||
QY 7441 CCCTTCTGACTTCCGCTTCCGACGAGCTGACAAAGGATCCGAGTGTGAGTCTGCTTCTCT 7500
DB |||||
QY 7442 CCCTTCTGACTTCCGCTTCCGACGAGCTGACAAAGGATCCGAGTGTGAGTCTGCTTCTCT 7501
DB |||||
QY 7501 CCATGCCCCCTTGAAGGGGACCGCGGACCGCGATCTTCAGGACGCTTGTGCTA 7560
DB |||||
QY 7502 CCATGCCCCCTTGAAGGGGACCGCGGACCGCGATCTTCAGGACGCTTGTGCTA 7561
DB |||||
QY 7561 CCGTGTGAGGAGGCTTGTGAGATCTGCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7620
DB |||||
QY 7562 CCGTGTGAGGAGGCTTGTGAGATCTGCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7621
DB |||||
QY 7621 CGGCGCTTGTACCGCTTCCGCTGCGGAGGAAAGTAAAGCTGCCCATCAACCGCTTGTGACCA 7680
DB |||||

Db	3610	GGGCGTTGGTTGACACCCAGATGCTTGGTCCACTACCCATACAGGCTTTGGCACTACCCC	3669
Qy	2199	TGCACCTCAATTTTTCATCTTTAAAGTTTAGGATGTATGGGGGGCTGGAGCACAGG	2258
Db	3670	TGCACCTCAATTTTACCATCTTCBAGGTTAGGATGTACGTGGGGGAGTGGAGCACAGG	3729
Qy	2259	CTCAATGCCGATAGCAATTGGACTCGAGGAGAGCGCTGTAACTTGGAGGACAGGGAATGG	2318
Db	3730	CTCGAAGCCGATGCAATTGGACTCGAGGAGAGCGCTGTAACTTGGAGGACAGGGAACAGA	3789
Qy	2319	TCAGAACTCAGCCCGCTCTGCTGTCTPACACAGAGTGGCAGATACCTGCCCCTGTGCTTTC	2378
Db	3790	TCAGAGCTTAGCCCGCTGCTGCTGCTACACAGGAGTGGCAGGATTTGCCCCTGTTCCTTC	3849
Qy	2379	ACCACCTTACCGGCTTTATCCACTGGTTTGATCCATCTCCATCAGAACTCGTGGACGTG	2438
Db	3850	ACCACCTTACCGGCTCTGCTCACTGGTTTGATCCATCTCCATCAGAACTCGTGGACGTG	3909
Qy	2439	CAATACCTGTACGGTGTAGGCTCAGCGTTTGTCTCTTTGCAATCAAAATGGAGTACATC	2498
Db	3910	CAATACCTGTACGGTGTAGGCTCAGCGGTTCTCTCTTTGCAATCAAAATGGAGTATGTC	3969
Qy	2499	CTGTGTCTTTCTCTCTCGCAGACGCGCGCTGTGTCCTGCTGTGTCGATGATGCTG	2558
Db	3970	CTGTGTCTTTCTCTCTCGCAGACGCGCGCTGTGTCCTGCTGTGTCGATGATGCTG	4029
Qy	2559	CTGATAGCCAGGTGAGGCGGCTTAGAGAACTTGGTGGTCTTCAATCGGCGCTCGCTG	2618
Db	4030	CTGATAGCTCAAGCTGAGGCGGCTTAGAGAACTTGGTGGTCTTCAATCGGCGCTCGCTG	4089
Qy	2619	GCCGAGGGCATGTATTTCTCTCTTTGTGTGTTCTTGGCGGCTCGGTACATTAAG	2678
Db	4090	GCCGAGGGCATGTATTTCTCTCTTTGTGTGTTCTTGGCGGCTCGGTACATTAAG	4149
Qy	2679	GGCAGGCTGGCTCTCGGGGGGCTATGCTTTTATGGCGTATGGCGCTGCTCTGCTC	2738
Db	4150	GGCAGGCTGGCTCTCGGGGGGCTATGCTTTTATGGCGTATGGCGCTGCTCTGCTC	4209
Qy	2739	CTACTGCGCTTACCAACACGAGCTTACGCTTGGACCGGAGATGGCTGCATCGTGGGG	2798
Db	4210	CTGCTGCGCTTACCAACACGAGCATACGCCATGGACCGGAGATGGCAGCATCGTGGGA	4269
Qy	2799	GGTGGCGTTCTGTAGTCTGGTATTTCTTGACCTGTGCAACAATCAAAAGTGTTCCT	2858
Db	4270	GGGCGGTTTCTGTAGTCTGGTATTTCTTGACCTGTGCAACAATCAAAAGTGTTCCT	4329
Qy	2859	ACTAGGCTCATATGGTGGTTACAAATCTTTATCAGCAGACCGAGGCGACATGCAAGT	2918
Db	4330	GCTAGGCTCATATGGTGGTTACAAATCTTTATCAGCAGGCGAGGCGACACTTGCAGT	4389
Qy	2919	TGGGTCCCCCTCAACGTTGGGGAGGCGGATGCCATCATCTCTTCAAGTGTGGG	2978
Db	4390	TGGATCCCCCTCAACGTTGGGGGGCGCGATGCCGTCACTCTCTCAGTGGCG	4449
Qy	2979	GTTTCATCCAGTTAATTTTGTACATCAACAACTCTGCTGGCAATCTGGGCCCGCTC	3038
Db	4450	ATCCACCCAGCTAATCTTTACCATCACAAATCTTGTCTGCCCATATCTCGTCCACTC	4509
Qy	3039	ATGGTCTCCAGGCTGGCATTAACAGAGTGGCGTACTTGTGGCGGCTCAAAGGCTCAT	3098
Db	4510	ATGGTCTCCAGGCTGGTATTAACAAAGTGGCGTACTTGTGGCGGCGCAACGGGCTCAT	4569
Qy	3099	CGTGCATGCAATGTTAGTGCAGAAAGTGGCGGGGTCATTTATCCAAATGGTCTCATG	3158
Db	4570	CGTGCATGCAATGTTAGTGCAGAAAGTGGTGGGGTCAATTTATGCCAAATGGCTCTCATG	4629
Qy	3159	AAGTGGGCGGCTGACAGTACGTAGTTTATPAACATCTTACCCCATCGGGGACTGG	3218
Db	4630	AAGTGGGCGGCTGACAGTACGTAGTTTATATGACCATCTCACCCCATCGGGGACTGG	4689
Qy	3219	GCCACAGCGGCGCTACAGACCTTGGGTGGCGGTAGACCGGTGTCTTCTCCGCGATG	3278
Db	4690	GCCACAGCGGCGCTACAGACCTTGGGTGGCGGTAGACCGGTGTCTTCTCTGATATG	4749

QY	3279	GAGACAAAGGCTATCACTGCTGGGAGCAGACACCGCTCGTGTGGGAGCATCATCTTGGGT	3339
Db	4750	GAGACAAAGGTTATCACTGCTGGGAGCAGACACCGCGCGTGTGGGACATCATCTTGGGC	4809
QY	3339	CTACCCGCTCTCGCCGGAAGGGGAAGAGATATTTTGGGACCGGCTGATAGTCTCGAA	3398
Db	4810	CTGCCCGTCTCGCCCGCAGGGGAGGAGATCATCTGGGACCGGAGACACGCTTTGAA	4869
QY	3399	GGGCAAGGCTGGGAGCTCTTTGGCCCAATCACGGCCCTACTCCCAACAAACCGCGGGCGTA	3458
Db	4870	GGGCAAGGCTGGGAGCTCTCGCGCCATTACGGCCCTACTCCCAACAGACCGCGAGGCTA	4929
QY	3459	CTTGGTTGCATCATCACTAGCGCTCACAGGCGCGGAGAGAGACCAAGTCTGAGAGGGAGT	3518
Db	4930	CTTGGCTCATCATCACTAGCGCTCACAGGCGCGGACAGGACCAAGTCTGAGGGGAGGTC	4989
QY	3519	CAAGTGGTTTCTACCGCAACAATCTTTCTGGCGACCTGCATCAACGGCTGTCTCG	3578
Db	4990	CAAGTGGTCTCCACCGCAACAATCTTTCTGGCGACCTGCCTCATGGCGTGTCTGG	504
QY	3579	ACTGTCTACCATGGCGTGGCTCGAGACCGTAGCGGTCGAAAAGTCCAAATCAACCCAA	3638
Db	5050	ACTGTCTATCATATGTGGCGGCTCAAGACCCCTTTCGGCGCCAAAGGGCCCAATCAACCCAA	5109
QY	3639	ATGTACACCAATGTAGACCTTGGACCTCGTGGGTGGCAGCGGCCCGCGGGCGCGCTCC	3698
Db	5110	ATGTACACCAATGTGACACGAGCCTGCTGGGTGGCAAGCGCCCGGGGCGCGCTTC	5169
QY	3699	ATGACACCATGCACTGTGGAGCTCGGACCTTTCGACTTGTGTACAGGACATGCTGATGT	3758
Db	5170	TTGACACCATGCACTCGGCGAGCTCGGACCTTTCGTTGGTACAGGAGCATGCCGATGT	5229
QY	3759	ATTCCGGTGGCGCGGAGCGCACAGAGGGAGTCTACTCTCCCGCAGCGCCGCTCC	3818
Db	5230	ATTCCGGTGGCGCGGAGCGCACAGAGGGAGGCTACTCTCCCGCAGGCGCGCTCC	5289
QY	3819	TACTGAAAGCTCCTCGGTTGGTTCATTGCTTTGCCCTTCGGGGCAGCTCGTGGGCGTC	3878
Db	5290	TACTGAAAGCTCCTCGGCGGTTCACCTGCTCTGCTCGCCCTGGGGCAGCGTGTGGGCAT	5349
QY	3879	TTCCGGGCTGCTGTGTGACACCGGGGGTCCGCAAGCGGTGGACTTATACCCGTTGAG	3938
Db	5350	TTTCCGGCTGCCGTGTGCAACCGAGGGTTCGCAAGCGGTGGACTTTTACCCGCTCAG	5409
QY	3939	TCTATGAAACTACATGCGGCTTCGCGTCTTCACAGACAACCTCAACCCCGCGGTGA	3998
Db	5410	TCTATGAAACCCTATGCGCTCCCGGCTTCACGGAACAATCGTCCCTCCGSCCGTA	5469
QY	3999	CCGAGACATTCGAAGTGGCATACTGTGCAAGCTCCTACTGGCAGCGGCAAGAGCACCAA	4058
Db	5470	CCGAGACATTCGAAGTGGCCATCTACACGCCCTACTGCTAGCGGCAAGAGCACTAAG	5529
QY	4059	GTCCGGCTCGTATGAGCCCAAGGTTCAAGGTCTCTCTCTCTGAAACCCGCTCCGTC	4118
Db	5530	GTCCGGCTCGTATGACGCCCAAGGTTAAGTGTCTTGTCTTGAAACCCGCTCCGTC	5589
QY	4119	GCACCTTAGGTTTGGGGGTATATGTCCGAGGACACGGTATCGACCCCTAACATCAGA	4178
Db	5590	GCCACCTTAGGTTTGGGGGTATATGTCTAAGGCACATGTTATCGACCCCTAACATCAGA	5649
QY	4179	ACTGGGTAGGACCATTTACCGGGCGGCTCCATTAAGTACTCTCAACCTATGGGCAAGTTC	4238
Db	5650	ACGGGGTAGGACCATTCACCGGCTGCCCCANTCAGTACTCCACCTATGGCAAGTTT	5709
QY	4239	CTTGCCGAGGTTGCTTCTGGGGCGCCTATGACATCATATATGTGATGAGTGCAC	4298
Db	5710	CTTGCCGAGGTTGCTTCTGGGGCGGCTATGACATCATATATGTGATGAGTGCAC	5769
QY	4299	TCAACTGACTCGACTACCATCTTGGGCATCGGACAGTCTTGAGACCAAGCGGCT	4358
Db	5770	TCAACTGACTCGACCACTATCTGTGGCATTCGCAAGCTTGAGACCAAGCGGCT	5829

QY	4359	GGAGCGCGCTCGTGTGCTCTGGCAACCGCTACA	CTCCGGGATCGGTACCGTCCACAC	4411
Db	5830	GGAGCGGACTCGTGTCTCTCGCACCGCTACGGCT	CCGGATCGGTACCGTCCACAT	5989
QY	4419	CCCAATATCGAGGAATAGGCTGTCCAACTATGG	AGAGATCCCGTTCTATGGCAAGCC	4478
Db	5890	CCAAACATCGAGGAGTGCTCTGTCCAGCACT	GGAGAAATCCCGTTTTATGGCAAGCC	5949
QY	4479	ATCCCCATTTAGGCCATCAAGGGGGGAGGCA	ATCTATTTCTGCCATTCCAAGAGAAA	4538
Db	5950	ATCCCCATCGAGACCATCAAGGGGGGAGGCA	CTCATTTCTGCCATTCCAAGAGAAA	6009
QY	4339	TGTGACGAGCTCGCGCAAAAGCTGACAGGCT	TCGGAATGACGCTAGCATATTACCGG	4598
Db	6010	TGTGATGAGCTCGCGCGAAAGCTGTCCGAGCT	TCGCGCTCGCATATGCTGTAGCATATTACCGG	6069
QY	4599	GGCCTTGATGTGCGTCATACCGCTATCGGAG	CGCTGTTCGCTGGCAACAGAGCT	4588
Db	6070	GGCCTTGATGTATCGTCAATACCAACTAGCG	GAGACGCTCATATGCTGTAGCAACGAGCGCT	6129
QY	4659	CTAATGACGGCTTTACCGCGCATTTTGACT	CTAGTGATCGACTCCAATACATGTGTCAAC	4718
Db	6130	CTAATGACGGCTTTACCGCGCATTTTGACT	CTAGTGATCGACTCCAATACATGTGTCAAC	6189
QY	4719	CAGACAGTGCATTCAGTTTGATCCCACTTCA	CCATTTGAGACGACGACCGTCCGCCAA	4778
Db	6190	CAGACAGTGCATTCAGTTTGATCCCACTTCA	CCATTTGAGACGACGACCGTCCGCCAA	6249
QY	4779	GACCGGTGTCCGCTCGCAACGGGAGGTAGA	CTGAGACTCGCAGGGGTAGAGTGGCATCTAC	4838
Db	6250	GACCGGTGTCTACGCTCGCAGCGGAGGAG	CTGTTAGGGGACGAGTGGGCATTTTAC	6309
QY	4839	AGGTTTGTGACTCCAGGAGAAACGGCCCT	CCGGGCATGTTCCGATTTCTCGGTCCCTGTGTGAG	4898
Db	6310	AGGTTTGTGACTCCAGGAGAAACGGCCCT	CCGGGCATGTTCCGATTTCTCGGTCCCTGTGTGAG	6369
QY	4899	TGCTATGACGGGGCTGTGTTGTGATGAGCT	ACGCCGCTGAGACCTCGGTTAGGTTG	4958
Db	6370	TGCTATGACGGGGCTGTGTTGTGATGAGCT	ACGCCGCTGAGACCTCGGTTAGGTTG	6429
QY	4959	CGGCTTACCTAAATACACGAGGTTCCCGT	CTGCCCAGACCATCTGGAGTTCTGGGAG	5018
Db	6430	CGGCTTACCTAAATACACGAGGTTCCCGT	CTGCCCAGACCATCTGGAGTTCTGGGAG	6489
QY	5019	AGCGCTTTACAGGCTTCACCCATAGATGCC	ACATTTCTGTGCCAGACTAAACAGGCA	5078
Db	6490	AGCGCTTTACAGGCTTCACCCATAGATGCC	ACATTTCTGTGCCAGACTAAACAGGCA	6549
QY	5079	GGAGACAACTTTCTTACCTGGTGGCATATCA	AGCTACAGTGTGGGCCAGGGCTCAAGCT	5138
Db	6550	GGAGACAACTTTCTTACCTGGTGGCATATCA	AGCTACAGTGTGGGCCAGGGCTCAAGCT	6609
QY	5139	CCACTCCATCGTGGGACCAATGTGAAGTGT	CTCATAGGCTGAAACTCAGCTGTCAC	5198
Db	6610	CCACTCCATCGTGGGACCAATGTGAAGTGT	CTCATAGGCTGAAACTCAGCTGTCAC	6669
QY	5199	GGGCCAAACCCCTGTGTATAGCTTAGGAG	CGCTCCAAAATAGAGGTCACTCTCACACAC	5258
Db	6670	GGGCCAAACCCCTGTGTATAGCTTAGGAG	CGCTCCAAAATAGAGGTCACTCTCACACAC	6729
QY	5259	CCCATATCAATATACATGATGGCATGATGT	CGGCTGACCTGGAGGTCTGCTACGAGCAC	5318
Db	6730	CCCATATCAATATACATGATGGCATGATGT	CGGCTGACCTGGAGGTCTGCTACGAGCAC	6789
QY	5319	TGGGTGCTGTTAGCGGAGTCTCTGACGCTT	TGGCGGCATACTGCTGTAGCAACGAGT	5378
Db	6790	TGGGTGCTGTTAGCGGAGTCTCTGACGCTT	TGGCGGCATACTGCTGTAGCAACGAGT	6849
QY	5379	GTGCTCATTTGGCGAGGATCATCTTCTCGG	GGAAGCCAGCTCTCGTTCCGACAGGAA	5438
Db	6850	GTGCTCATTTGGCGAGGATCATCTTCTCGG	GGAAGCCAGCTCTCGTTCCGACAGGAA	6909
QY	5439	GTCCTTACAGGAGTTTCGATGAGATGGA	AGAGTGTGCTTCACAACTCTCTTACATCGAG	5498

D5	6910	GTCTTTTACCGGAGTTTCGATGAGATGGAGAGTGGCCTTCAACCTCCCTTACATCGAA	6966
QY	5499	CAGGGAATGCACTCGCGAGCAATTCAAGCAAAAGCGCTCGGGTGTGTGCAAAAGGCC	5558
D5	6970	CAGGGAATGCACTCGCGAACAATTCAACAGAGGCAATCGGTTGCTGCAACAGCC	7029
QY	5559	ACCAAGCAAGGAGAGGCTGTGTCTCGTGTGTGGAGTCCAAGTGGCGAGCCCTTGAGACC	5618
D5	7030	ACCAAGCAAGCGGAGGCTGTGTCTCGGTGTGTGAATCCAAGTGGCGGAACCTTCGAAGCC	7089
QY	5619	TTCTGGCGGAAGCACATGTGGAATTTTCATCAGCGGAATACAGTACCTAGCAGGCTTATCC	5678
D5	7090	TTCTGGCGGAAGCAATGTGGAATTTTCATCAGCGGAATACATATTTAGCAGGCTTGTCC	7149
QY	5679	ACTCTGCTGGAACCCCGCATAGCATCATTCATGGCATTTACAGCTTCTATCACTAGC	5738
D5	7150	ACTCTGCTGGCAACCCCGCATAGCATCACTGATGCAATTCACAGCCTCTATCACAGC	7209
QY	5739	CCGCTGACGACCCAAACACCTCTCTGTTTAAACATCTTGGGGGGAATGGGTGGCTGCCAA	5798
D5	7210	CCGCTGACCAACCCAAACATACCTCTCTGTTTAAACATCTCGGGGGAATGGGTGGCGCCAA	7269
QY	5799	CTCGCTCTCCACAGCGCTGTGTCAAGCTTTCGTGGGCGCGGCATCCCGCAGCGGCTGTT	5858
D5	7270	CTTGTCTCTCCAGCGCTGCTTCTGTCTTCGTAGGCGCGGCATCGCTGGAGCGGCTGTT	7329
QY	5859	GGCAGCAATGAGCTTGGAAAGTGCTCGHGAACATTTGGCGGGCTATGGGCGAGGGTA	5918
D5	7330	GGCAGCATAGGCTTGGGAAGTGTCTGTGGATATTTGGCAGTTATGAGCAGCGCTG	7389
QY	5919	GCGGGCGCATCGTGGCTTAAAGTTCATAGCGGCGAGGTGCCCTCCACCGAGGACCTG	5978
D5	7390	GCAGGCGCGTCTGTGGCTTAAAGTTCATAGCGGCGAGATGCCCTCCACCGAGGACCTG	7449
QY	5979	GTCAACTTACTCCCTGCATCCCTCTCTCTGTGTGCCCTGTGTCTGGGTGTGTGCGCA	6038
D5	7450	GTTTAACTATCCCTGCTATCCTCTCCCTGGCGCCTTAGTGTGTGGGGTGTGTGCGCA	7509
QY	6039	GCAATACTGTGTGGCAGCTGGGCGCGGAGAGGGGCTGTGTGAGTGAATGAACGGCTG	6098
D5	7510	GCGATATGTGTGGCAGCTGTGGCGCCAGGAGGGGGCTGTGTGAGTGAATGAACGGCTG	7569
QY	6099	ATAGGCTGCTTCGCGGGTAACACGCTTCCTTACGCACTNTGTGCTTGAGCGAC	6158
D5	7570	ATAGGCTGCTTCGCGGGTAACACGCTTCCTTACGCACTNTGTGCTTGAGCGAC	7629
QY	6159	GCTGCGACAGTGTCACTCAGATCCCTCTCTAGCCTTACCATCACTCAACTGTGTAAGCG	6218
D5	7630	GCTGCGACAGTGTCACTCAGATCCCTCTCTAGTCTTACCATCACTCAGTGTGTGAAGG	7689
QY	6219	CTCCACAGTGGATTAATGAGGACTGCTCTTAAGCCATGTCTCGGCTGTGGCTTAAGGAT	6278
D5	7690	CTTTCACAGTGGATCAACAGGACTGCTCCAGCCCATGCTCGGCTCGTGGCTAAGAT	7749
QY	6279	GTTTGGATTGGATATGCAAGGTGTGACTGACTTCAAGACCTGGCTCCAGTCCAACTC	6338
D5	7750	GTTTGGATTGGATATGCAAGGTGTGACTGATTTCAAGACCTGGCTCCAGTCCAACTC	7809
QY	6339	CTCCCGGCTTACCGGAGTCCCTTTCTCTGTATGCCAAAGCGGGTACAAAGGAGTCTGG	6398
D5	7810	CTCCCGGATGCGGGAGTCCCTTCTCTCATGTCACTGCTGGGTACAGGAGTCTGG	7869
QY	6399	CGGGGGGACGGCATCATGAAACCACTGCCCATGCGGAGACAGATCCCGGACATGTC	6458
D5	7870	CGGGGGGACGGCATCATGAAACCACTGCCCATGCGGAGACAGATCAACCGGACATGTG	7929
QY	6459	AAAAAGGTTCCATGAGGATCTGAGGCGCTAGAACTGCGAGCAACAGCTGGCACGGAAG	6518
D5	7930	AAAAAGGTTCCATGAGGATCTGAGGCGCTAGAACCTGTGTATCAACAGCTGGCATGAA	7989
QY	6519	TTCCCATCAACGCATACACCACGGGACCTTGCAACCTCCCGCGCGCCCAACTATTC	6578

Df	7990	TTCCCAATTAACCGGTACACACACGGGCCCCGACGCGCTCCCGGCGCCAAATTAATCT	8049
Qy	6579	AGGCGCTATGGCGGTGGCTGCTCAGGAGTACGTGGAGGTATACGCTGTGGGGATTC	6638
Df	8050	AGGCGCTGTGGCGGTGGCTGCTCAGGAGTACGTGGAGGTATACGCTGTGGGGATTC	8109
Qy	6639	CACATCGTACGGGCGATGACCACTGACCAACGTAAAGTGCCTATGCGAGGTTCCGCGCCCC	6698
Df	8110	CACATCGTACGGGCGATGACCACTGACCAACGTAAAGTGCCTATGCGAGGTTCCGCGCCCC	8169
Qy	6699	GAATTCCTACGAGAGTGGATGGATGGCTGGCTGACAGTACGCTCCGCGGTGGAACCT	6758
Df	8170	GAATTCCTACGAGAGTGGATGGATGGCTGGCTGACAGTACGCTCCGCGGTGGAACCT	8229
Qy	6759	CTTCTACGGGAGGAGCTCAAGCTCCAGGTCCGGCTCAACCAATATCTGCTGGGTCCGAG	6818
Df	8230	CTTCTACGGGAGGAGCTCAAGCTCCAGGTCCGGCTCAACCAATATCTGCTGGGTCCGAG	8289
Qy	6819	CTCCATCGGAGCCGGAACCGGAGCTAACAGTGTCTTCTCCATGCTCACCGATCCCTCC	6878
Df	8290	CTCCATCGGAGCCGGAACCGGAGCTAACAGTGTCTTCTCCATGCTCACCGATCCCTCC	8349
Qy	6879	CACATTACAGCAGAGACGGCTAAGCTAGGCTGGCTAGAGGTCTCCCGCTCTTTAGCC	6938
Df	8350	CACATTACAGCAGAGACGGCTAAGCTAGGCTGGCTAGAGGTCTCCCGCTCTTTAGCC	8409
Qy	6939	AGCTCATCAGCTAGCCAGTGTCTGCGCTCTTTTGAAGGGAGATGCACTACCCACCAT	6998
Df	8410	AGCTCATCAGCTAGCCAGTGTCTGCGCTCTTTTGAAGGGAGATGCACTACCCACCAT	8469
Qy	6999	GACTCCCGGAGCGTGAACCTCATCGAGGCCAACCTCTTTGGCGGCGAGAGATGGCGGA	7058
Df	8470	GACTCCCGGAGCGTGAACCTCATCGAGGCCAACCTCTTTGGCGGCGAGAGATGGCGGA	8529
Qy	7059	AACATCACTCCGCTGGAGTACAGAGATAAGTAGTAATCTGGACTCTTTGCAACCGCTT	7118
Df	8530	AACATCACTCCGCTGGAGTACAGAGATAAGTAGTAATCTTTGGACTCTTTGCAACCGCTT	8589
Qy	7119	CACGCGAGGGGGATGAGAGGAGATATCCGTCGCGGGAGATCTCTCGGAAATCCAGG	7178
Df	8590	CACGCGAGGGGGATGAGAGGAGATATCCGTCGCGGGAGATCTCTCGGAAATCCAGG	8649
Qy	7179	AAGTTCCCTCAGCGTTCGCCATATGGGCAACCGCGGACTACAATCTCCACTGCTAGAG	7238
Df	8650	AAATTCCTCAGCGTTCGCCATATGGGCAACCGCGGACTACAATCTCCACTGCTAGAG	8709
Qy	7239	TCCTGGAAGGACCGGACTACGTCCTCCGATGGTACACGATGCGCAATTCGCACTACC	7298
Df	8710	TCCTGGAAGGACCGGACTACGTCCTCCGATGGTACACGATGCGCAATTCGCACTACC	8769
Qy	7299	AAGGCTCCTCCAAATACCACTCCAGGAGAAAGAGGACGGTTGCTCCGACAAATCCAA	7358
Df	8770	AAGGCTCCTCCAAATACCACTCCAGGAGAAAGAGGACGGTTGCTCCGACAAATCCAA	8829
Qy	7359	GTGTCCTTCGCTTGGCGAGCTCGCACTAAGACCTTCGGTAGCTCCGGATTCGTGGCC	7418
Df	8830	GTGTCCTTCGCTTGGCGAGCTCGCACTAAGACCTTCGGTAGCTCCGGATTCGTGGCC	8899
Qy	7419	GTTCATAGCGGACGGGACCGCTCTCTGACCTGGCTCCGACGAGGGTGACAAAGGA	7478
Df	8890	GTTCATAGCGGACGGGACCGCTCTCTGACCTGGCTCCGACGAGGGTGACAAAGGA	8949
Qy	7479	TCCGAGTGTAGTCTACTCTCTCCATGCCCCCTTTGAAGGGGAGCGGGGACCCCGAT	7538
Df	8950	TCCGAGTGTAGTCTACTCTCTCCATGCCCCCTTTGAAGGGGAGCGGGGATCCCGAT	9009
Qy	7539	CTCAGCGAGGGTCTTGCTTACCGTGTAGGAGGCTATGAGGATGTGCTGTGCTGC	7598
Df	9010	CTCAGCGAGGGTCTTGCTTACCGTGTAGGAGGCTATGAGGATGTGCTGTGCTGC	9069
Qy	7599	TCAATGTCTTACGTGACAGGGGCTGTATGATCAGCCATCCGCTGGGAGGAAAGTAA	7658
Df	9070	TCGATGTCTTACATGACAGGGGCTGTATGATCAGCCATCCGCTGGGAGGAAACCAAG	9129
Qy	7659	CTGCCATGACACCGTGTGAGAACTCTTTGCTGGTGTCCACACACATGCTGTAGCCACA	7718
Df	9130	CTGCCATGACACCGTGTGAGAACTCTTTGCTGGTGTCCACACACATGCTGTAGCCACA	9189
Qy	7719	ACATCCCGCAGCGCAAGCTCCGCGAGAAAGGTACCTTTGACAGATGCAAGTCTCTG	7778
Df	9190	ACATCCCGCAGCGCAAGCTCCGCGAGAAAGGTACCTTTGACAGATGCAAGTCTCTG	9249
Qy	7779	GATGATCATTTACCGGAGCGTATCTCAGAGCATGAAAGCGAAGCGTCCACAGTTAAGCT	7838
Df	9250	GATGATCATTTACCGGAGCGTATCTCAGAGCATGAAAGCGAAGCGTCCACAGTTAAGCT	9309
Qy	7839	AAGCTTCTATCTATAGAGGAGGCTGCAAGCTGAGCGCCCACTTCCGCGCAATCCAA	7898
Df	9310	AAGCTTCTATCTGAGAGGAGCTGTAAGCTGAGCGCCCACTTCCGCGCAATCCAA	9369
Qy	7899	TTTGGCTATGGGCAAGAGCGTCCGGAACCTATCCAGCAGGGCGCTTAAACCATCCGC	7958
Df	9370	TTTGGCTATGGGCAAGAGCGTCCGGAACCTATCCAGCAGGGCGCTTAAACCATCCGC	9429
Qy	7959	TCGCTGGGAGGAGCTGCTGAGAGACATGAAACACCAATTTGACACCAATCATGGCA	8018
Df	9430	TCGCTGGGAGGAGCTGCTGAGAGACATGAAACACCAATTTGACACCAATCATGGCA	9489
Qy	8019	AAAAGTGAAGCTTTCTGCTCCAAACAGAGAGGAGCGCGAAGCGAGCTCCGCTTATC	8078
Df	9490	AAAAGTGAAGCTTTCTGCTCCAAACAGAGAGGAGCGCGAAGCGAGCTCCGCTTATC	9549
Qy	8079	GTATTCACAGCTGGGAGTTCGTATGAGAGATGGCCCTTTACGATGGTCTCC	8138
Df	9550	GTATTCACAGCTGGGAGTTCGTATGAGAGATGGCCCTTTACGATGGTCTCC	9609
Qy	8139	ACCTTCTCAGCGCTGATGGCTCTCATAGGATTTCAATACCTCCCGAAGCGGG	8198
Df	9610	ACCTTCTCAGCGCTGATGGCTCTCATAGGATTTCAATACCTCCCGAAGCGGG	9669
Qy	8199	GTGAGTTCGTGTGTAATCTCGAAATCAAGAAATGCTTATGGCTTCTCATATGAC	8258
Df	9670	GTGAGTTCGTGTGTAATCTCGAAATCAAGAAATGCTTATGGCTTCTCATATGAC	9729
Qy	8259	ACCGCTGTTTTGACTCAACGCTCACTGAGATGACATTCGTGTGAGGAGTCAATTTAC	8318
Df	9730	ACCGCTGTTTTGACTCAACGCTCACTGAGATGACATTCGTGTGAGGAGTCAATTTAC	9789
Qy	8319	CAATGTTGACTTGGCCCCCGGAGGAGACAGGCAATAGGTCGCTCAGAGCGGCTT	8378
Df	9790	CAATGTTGACTTGGCCCCCGGAGGAGACAGGCAATAGGTCGCTCAGAGCGGCTT	9849
Qy	8379	TACATCGGGGCTCCCTGACTAACTCAAAAGGCGAGAACTGCGGTTATGCGCGTTCGCG	8438
Df	9850	TACATCGGGGCTCCCTGACTAACTCAAAAGGCGAGAACTGCGGTTATGCGCGTTCGCG	9909
Qy	8439	GCAAGTGCCTGTGAGTACGATGCTGGGTAATACCTCTCATATGTTACTTGAAGGCACT	8498
Df	9910	GCAAGTGCCTGTGAGTACGATGCTGGGTAATACCTCTCATATGTTACTTGAAGGCACT	9969
Qy	8499	GCAAGCTGTGAGTCAAAAGCTCCAGGCTGACAGTGTCTGAGACGAGGAGGACCTT	8558
Df	9970	GCAAGCTGTGAGTCAAAAGCTCCAGGCTGACAGTGTCTGAGACGAGGAGGACCTT	10029
Qy	8559	GTGCTTATCTGTGAAAGCGGGGAAACCCAGAGGAGTGGCGGCGCTTACGAGCTTCAAG	8618
Df	10030	GTGCTTATCTGTGAAAGCGGGGAAACCCAGAGGAGTGGCGGCGCTTACGAGCTTCAAG	10089
Qy	8619	GAGGCTATGATAGGTTATTCGCGCCCCCGGGGATCCGCGCCCAACAGAAATACGACCTG	8678
Df	10090	GAGGCTATGATAGGTTATTCGCGCCCCCGGGGATCCGCGCCCAACAGAAATACGACCTG	10149
Qy	8679	GAGCTGATACATCATGTTCTCTCAATGTCTAGTGGCGGACGATGCTGSCAAAGG	8738
Df	10150	GAGTGTGATACATCATGTTCTCTCAATGTCTAGTGGCGGACGATGCTGSCAAAGG	10209

```
QY 8739 GTATACTACTCACCGCTGACCCACACCCCTTGTACAGCGGCTGCGTGGAGACAGCT 8798
Db 10210 GGTACTATCTCACCGCTGACCCACACCCCTTGTACAGCGGCTGCGTGGAGACAGCT 10269
QY 8799 AGACACATCCCAATCAACTCTTGGCTAGCAATATCATCATGTATGGGCCACCCCTATGG 8858
Db 10270 AGACACATCCCAATCAACTCTTGGCTAGCAATATCATCATGTATGGGCCACCCCTATGG 10329
QY 8859 GCAAGGATGATCTGATGACTCACTTTTCTCCATCTCTAGCTCAAGAGCAACTTGA 8918
Db 10330 GCAAGGATGATCTGATGACTCACTTTTCTCCATCTCTAGCTCAAGAGCAACTTGA 10389
QY 8919 AAAGCCCTGGATGTGATGACTCACTTTTCTCCATCTCTAGCTCAAGAGCAACTTGA 8978
Db 10390 AAAGCCCTGGATGTGATGACTCACTTTTCTCCATCTCTAGCTCAAGAGCAACTTGA 10449
QY 8979 CAGATCATGAACGACTCAATGCTCTTAGCGATTTACTCACTCAAGTACTCTCCAGT 9038
Db 10450 CAGATCATGAACGACTCAATGCTCTTAGCGATTTACTCACTCAAGTACTCTCCAGT 10509
QY 9039 GAGATCAATAGGCTGCTTCATGCTTCAGGAACTTGGGTACCACTTGGAACTTGG 9098
Db 10510 GAGATCAATAGGCTGCTTCATGCTTCAGGAACTTGGGTACCACTTGGAACTTGG 10569
QY 9099 AGACATCGGGCCAGAAAGTGTCCGCTAGCTACTCTCCAGGGGGAGGGCCGCCACT 9158
Db 10570 AGACATCGGGCCAGAAAGTGTCCGCTAGCTACTCTCCAGGGGGAGGGCCGCCACT 10629
QY 9159 TGTGCGCATACCTCTTTAACTGGGAGTAAAGGACCAAGCTTAACTCACTCCAACTCCG 9218
Db 10630 TGTGCGCATACCTCTTTAACTGGGAGTAAAGGACCAAGCTTAACTCACTCCAACTCCG 10689
QY 9219 GCGCGTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9278
Db 10690 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10749
QY 9279 TATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9338
Db 10750 TATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10809
QY 9339 GTAGGGTAGGATTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9398
Db 10810 GTAGGGTAGGATTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10869
QY 9399 TTAAGCC-----ATTTCCTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 9448
Db 10870 AATAGGCCATCTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 10929
QY 9449 TTTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 9508
Db 10930 TTTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 10989
QY 9509 CTTAGCCCTAGTCAGGCTAGCTGGAAGTCCGTGAGCGCATCTAGTCAGAGAGTGC 9568
Db 10990 CTTAGCCCTAGTCAGGCTAGCTGGAAGTCCGTGAGCGCATCTAGTCAGAGAGTGC 11049
QY 9569 TGAATAGTGCCTCTCTGCGATCATGT 9595
Db 11050 TGAATAGTGCCTCTCTGCGATCAAT 11076
```

RESULT 9

```
US-09-539-601-25
; Sequence 25, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
```

```
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 11076
; TYPE: DNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(341)
; OTHER INFORMATION: construct I389/Core-3'/5.1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342)..(1193)
; OTHER INFORMATION: hepatitis C virus core - neomycin
; OTHER INFORMATION: phosphotransferase fusion protein
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (1202)..(1812)
; OTHER INFORMATION: internal ribosome entry site from
; OTHER INFORMATION: encephalomyocarditis virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1813)..(10845)
; OTHER INFORMATION: hepatitis C virus polyprotein from core to
; OTHER INFORMATION: nonstructural protein NS5B; carries cell
; OTHER INFORMATION: culture-adaptive mutations of clone 5.1
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (10846)..(11076)
; US-09-539-601-25
```

Query March 82.3%; Score 7901; DB 4; Length 11076;
Best Local Similarity 90.9%; Pred. No. 0;
Matches 8422; Conservative 0; Mismatches 835; Indels 10; Gaps 1;

```
QY 339 ACCATGAGCAGCAATCTTAACCTCAAAGAAACCAAAAGTAAACACCAACCGCGGCCA 398
Db 1810 ACCATGAGCAGCAATCTTAACCTCAAAGAAACCAAAAGTAAACACCAACCGCGGCCA 1869
QY 399 CAGGACGTCAGTTCGCGGGGGTGTGTCAGATCGTTGGTGGAGTTTACCTGTTCCGCGCC 458
Db 1870 CAGGACGTCAGTTCGCGGGGGTGTGTCAGATCGTTGGTGGAGTTTACCTGTTCCGCGCC 1929
QY 459 AGGGGCCCGCGAGTTGGTGTGCGCGGAGCTAGGAGGCTTCCGAGCGTCCGACCTCGT 518
Db 1930 AGGGGCCCGCGAGTTGGTGTGCGCGGAGCTAGGAGGCTTCCGAGCGTCCGACCTCGT 1989
QY 519 GGAAGGCGCAAACTATCCCAAGGCTCGCGGACCCGAGGGGCGAGGGCTTGGGCTCAGGCC 578
Db 1990 GGAAGGCGCAAACTATCCCAAGGCTCGCGGACCCGAGGGGCGAGGGCTTGGGCTCAGGCC 2049
QY 579 GGTACCTTTGGCCCTCTATGGAATGAGGCTTGGGTTGGGCGAGATGGCTCTCTGCA 638
Db 2050 GGTACCTTTGGCCCTCTATGGAATGAGGCTTGGGTTGGGCGAGATGGCTCTCTGCA 2109
QY 639 CCCCGCGCTCCCGGCTAGTTGGGGCCCAAGGAGGCTTGGGGTGGGCGAGATGGCTCTCTGCA 698
Db 2110 CCCCGCTCTCTGCGCTAGTTGGGGCCCAAGGAGGCTTGGGGTGGGCGAGATGGCTCTCTGCA 2169
QY 699 GGTAAAGTCAATGATACCTTATCATGCGGCTTCCGCGATCTCATGGGGTACATTCGCGCTC 758
Db 2170 GGTAAAGTCAATGATACCTTATCATGCGGCTTCCGCGATCTCATGGGGTACATTCGCGCTC 2229
QY 759 GTGGCGCCCGCTTGGGGCGGCTTCCAGGGCTTGGGCAACGGTGTCCGGTCTGGAG 818
Db 2230 GTGGCGCCCGCTTGGGGCGGCTTCCAGGGCTTGGGCAACGGTGTCCGGTCTGGAG 2289
QY 819 GAGCGGTGAACTATGCAACAGGAGACTTCCCGGTGCTCTTCTCTATCTCTCTCTTTC 878
Db 2290 GAGCGGTGAACTATGCAACAGGAGACTTCCCGGTGCTCTTCTCTATCTCTCTCTTTC 2349
QY 879 GCTCTGCTCTCTGTTTGTGACCATCCAGCTTCCGCTTATGAGTGGCGCAACGTGTCCGGG 938
```


Dd 2350 GCTTTGCTGCTCTGTTGACATCCAGCTTCCGCTTATGAGTGCAGCAAGTATCCGGA 2409
Qy 939 ATATACCATGTACAGAAAGCTGTCTCCAACTCAAGCATTTGTATGAGGACAGCGAGTG 998
Dd 2410 GTGTACCATGTACAGAAAGCTGTCTCCAAAGCATTTGTATGAGGACAGCGACATG 2469
Qy 999 ATCATGATATATCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1058
Dd 2470 ATCATGATATATCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2529
Qy 1059 GTAGGCTCACTCCCAAGCTCGCGGACAGGATGCCAGCTCCCACTACACATACGA 1118
Dd 2530 GTAGGCTCACTCCCAAGCTCGCGGACAGGATGCCAGCTCCCACTACACATACGA 2589
Qy 1119 CGCACGTCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1178
Dd 2590 CGCCATGTCGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2649
Qy 1179 CTCTGCGGATCTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1238
Dd 2650 CTCTGCGGATCTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2709
Qy 1239 ACAGTCAGGACTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1298
Dd 2710 ACAGTCAGGACTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2769
Qy 1299 TGGGATATGATGATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1358
Dd 2770 TGGGATATGATGATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2829
Qy 1359 ATCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1418
Dd 2830 ATCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2889
Qy 1419 GCCTACTATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1478
Dd 2890 GCCTACTATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2949
Qy 1479 GCGCTTACGCGGAGACCCACAGCGGAGGCTGCGCGGACACACACCTCCGCGGCTC 1538
Dd 2950 GCGCTTACGCGGAGACCCATATGCTGACAGGCGGACGATGCGGACACACCTCCGCGGAT 3009
Qy 1539 ACCTCCCTTTTCTCATCTGCGGCTCTCAGAAATCCAGCTTGTGATACCAACCGGACG 1598
Dd 3010 ACCTCCCTTTTCTCATCTGCGGCTCTCAGAAATCCAGCTTGTGATACCAACCGGACG 3069
Qy 1599 TGGCACATCAACAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1658
Dd 3070 TGGCACATCAACAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3129
Qy 1659 GCGCTGCTTACGACACAGATTTCACTGCTGCGGCTGCGGAGCGCATGCGGAGCTG 1718
Dd 3130 GCGCTGCTTACGACACAGATTTCACTGCTGCGGCTGCGGAGCGCATGCGGAGCTG 3189
Qy 1719 CGCCCCATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778
Dd 3190 AGCCCCATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3249
Qy 1779 TCGGATCAGAGGCTTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1838
Dd 3250 TCGGATCAGAGGCTTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3309
Qy 1839 TCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1898
Dd 3310 GCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3369
Qy 1899 GATGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1958
Dd 3370 GATGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3429
Qy 1959 AACCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2018
Dd 3430 AACCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3489

Qy 2019 TTCACTAAGACGTCGCGAGTCCCGGTGTAGTAAATCGGGGGGTCGTAACCGACCTTG 2078
Dd 3490 TTCACTAAGACGTCGCGAGTCCCGGTGTAGTAAATCGGGGGGTCGTAACCGACCTTG 3549
Qy 2079 ATCTGCCCCACGACGTCCTTCGCGAAGCAACCCGAGGCTACTTTACACAAATGTGGCTG 2138
Dd 3550 ACCTGCCCCACGACGTCCTTCGCGAAGCAACCCGAGGCTACTTTACACAAATGTGGCTG 3609
Qy 2139 GGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2198
Dd 3610 GGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3669
Qy 2199 TGCACTCTCAATTTTCCATCTTTAAGTTAGATGATGATGTTGGGGGCGTGGAGCACAGG 2258
Dd 3670 TGCACTCTCAATTTTCCATCTTTAAGTTAGATGATGATGTTGGGGGCGTGGAGCACAGG 3729
Qy 2259 CTCAATGCCCATGCAATTTGCACTCGAGGAGCGCTGTAACTTGGAGGACAGGATAGG 2318
Dd 3730 CTCAATGCCCATGCAATTTGCACTCGAGGAGCGCTGTAACTTGGAGGACAGGATAGG 3789
Qy 2319 TCAGAACTCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2378
Dd 3790 TCAGAACTCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3849
Qy 2379 ACCACCTACCGCTTATCCACTGCTTATCCACTGCTTATCCACTGCTTATCCACTGCTTATCCACTG 2438
Dd 3850 ACCACCTACCGCTTATCCACTGCTTATCCACTGCTTATCCACTGCTTATCCACTGCTTATCCACTG 3909
Qy 2439 CAATACTG 2498
Dd 3910 CAATACTG 3969
Qy 2499 CTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2558
Dd 3970 CTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4029
Qy 2559 CTGATAGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2618
Dd 4030 CTGATAGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4089
Qy 2619 GCGGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2678
Dd 4090 GCGGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4149
Qy 2679 GCGGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2738
Dd 4150 GCGGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4209
Qy 2739 CTACTGCGGTTACCAACAGGCTTACGCTTGGACCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2798
Dd 4210 CTACTGCGGTTACCAACAGGCTTACGCTTGGACCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4269
Qy 2799 GGTGCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2858
Dd 4270 GGTGCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4329
Qy 2859 ACTAGGCTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2918
Dd 4330 GCTAGGCTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4389
Qy 2919 TGGGTCCCCCCTCAACGTTTCGGGAGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2978
Dd 4390 TGGATCCCCCCTCAACGTTTCGGGAGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4449
Qy 2979 GTTCATCCAGGTTAATTTTTCATCATCAACAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3038
Dd 4450 ATCCACCCAGGCTAATTTTATCCATCACAAAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4509
Qy 3039 ATGCTGCTCAGGCTGGCATAAAGGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3098
Dd 4510 ATGCTGCTCAGGCTGGCATAAAGGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4569

QY 3099 CTTGATGATGTTAGTGCAGAAAGTCCGGGGGTCAATATGTCACAAATGCTCTCATG 3158
 Db 4570 CTTGATGATGATGCTGCTGCGAAGGTGCTGGGGTCAATATGTCACAAATGCTCTCATG 4629
 QY 3159 AAGCTGGGCGCGTGCAGAGTACATGATGTTATACCATCTTACCCCATGCGGGAAGTGG 3218
 Db 4630 AAGTTGGCGCGACTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 4689
 QY 3219 GCGCAAGGGGCTACAGACCTTGGCGTGGGGTAGAGCCGTGCTCTTCTCCGCCATG 3278
 Db 4690 GCGCAAGGGGCTACAGACCTTGGCGTGGGGTAGAGCCGTGCTCTTCTCTGATAG 4749
 QY 3279 GAGACAAAGGTATCATCACTGGGAGAGACACCGCTGCGTGTGGGACATCATCTTGGGT 3338
 Db 4750 GAGACAAAGGTATCATCACTGGGAGAGACACCGCGGCTGTGGGACATCATCTTGGGC 4809
 QY 3339 CTACCGCTTCGCGCGGAGAGGAGAGATATTTTGGGACCGCTGATATGCTCGAA 3398
 Db 4810 CTGCGCGCTTCGCGCGGAGAGGAGAGATATCATCTGGGACCGGACAGACGCTTGAA 4869
 QY 3399 GGGCAAGGGTGGCGACTCTGCGCCATCAAGGCTTACTCCCAACAAACGGGGCGTA 3458
 Db 4870 GGGCAAGGGTGGCGACTCTGCGCGCTATTAAGGCTTACTCCCAACAGAGCGGAGCTA 4929
 QY 3459 CTTGTTGATCATCATCACTAGCTTCAGCGCGGAGACAAAGCAGGTTCGAGGGAGGTT 3518
 Db 4930 CTTGTTGATCATCATCACTAGCTTCAGCGCGGAGACAAAGCAGGTTCGAGGGAGGTC 4989
 QY 3519 CAAGTGGTTTACCGCAACAACTTTCTTGGGAGCTTGTGATCAAGCGGCTGTGCTGG 3578
 Db 4990 CAAGTGGTTTCTCACCGCAACAACTTTCTTGGGAGCTTGTGATCAAGCGGCTGTGCTGG 5049
 QY 3579 ACTGTCTACCATGGCGCTGGTTCGAAGACCTTAGCGGTCGAAAGGTCCAAATACCCAA 3638
 Db 5050 ACTGTCTACCATGGCGCTGGTTCGAAGACCTTAGCGGTCGAAAGGTCCAAATACCCAA 5109
 QY 3639 ATGTACCAATATGATAGCTGATGATCTGTCGCTGTCAGGCGCGCGCGGCTGCTCC 3698
 Db 5110 ATGTACCAATATGATAGCTGATGATCTGTCGCTGTCAGGCGCGCGCGGCTGCTCC 5169
 QY 3699 ATGACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3758
 Db 5170 ATGACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5229
 QY 3759 ATTCGGGTGCGCGGAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3818
 Db 5230 ATTCGGGTGCGCGGAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5289
 QY 3819 TACCTGAAAGGCTCCTCGGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3878
 Db 5290 TACCTGAAAGGCTCCTCGGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5349
 QY 3879 ATTCGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3938
 Db 5350 ATTCGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5409
 QY 3939 TCTATGAAAGTACCATGCGGCTCTCGGCTCTTCAAGAGCACTCAACCCCGCGCTGTA 3998
 Db 5410 TCTATGAAAGTACCATGCGGCTCTCGGCTCTTCAAGAGCACTCAACCCCGCGCTGTA 5469
 QY 3999 CCGCAGACATTCAGATGAGCAATCTGACGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 4058
 Db 5470 CCGCAGACATTCAGATGAGCAATCTGACGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 5529
 QY 4059 GTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4118
 Db 5530 GTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5589
 QY 4119 GCCACCTTAGGTTTGGGCGGTATATGTCAGAGGACACGCTATGACCGCTTAACATCAGA 4178
 Db 5590 GCCACCTTAGGTTTGGGCGGTATATGTCAGAGGACACGCTATGACCGCTTAACATCAGA 5649
 QY 4179 ACTGGGTAAAGACCAATTACACCGGCGGCTCCATTAAGTACTCCACCTATGGCAAGTTC 4238

Db 5650 ATCGGGTAAAGACCATCACACCGGTCCCGCCCATCGTACTCCACTATGGCAAGTTT 5709
 QY 4239 CTTGCGGAGGTGGTGTCTGGGGGCGCTTANGACATCAATAATGATGATGATGATGATG 4298
 Db 5710 CTTGCGGAGGTGGTGTCTGGGGGCGCTTANGACATCAATAATGATGATGATGATGATG 5769
 QY 4299 TCAACTGATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4358
 Db 5770 TCAACTGATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5829
 QY 4359 GAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4418
 Db 5830 GAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5889
 QY 4419 CCAATATCGAGGAAATAGGCTGTCCAAATATGAGAGATGCCCTTCTATGGGAAAGCC 4478
 Db 5890 CCAATATCGAGGAAATAGGCTGTCCAAATATGAGAGATGCCCTTCTATGGGAAAGCC 5949
 QY 4479 ATCCCATTTAGGCGCATCAAGGGGGAGGAGCATCTCAATTTCTGCTCAATCCAGAGAA 4538
 Db 5950 ATCCCATTTAGGCGCATCAAGGGGGAGGAGCATCTCAATTTCTGCTCAATCCAGAGAA 6009
 QY 4539 TGTGACGAGCTCGCGCAAGCTGACAGGCTTCGACCTGAAAGCTGTAGCATATTTACCGG 4598
 Db 6010 TGTGACGAGCTCGCGCAAGCTGACAGGCTTCGACCTGAAAGCTGTAGCATATTTACCGG 6069
 QY 4599 GGCCTTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4658
 Db 6070 GGCCTTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6129
 QY 4659 CTAATGACGGGTTTACCGCGGATTTTACCGCGGATTTTACCGCGGATTTTACCGCGG 4718
 Db 6130 CTAATGACGGGTTTACCGCGGATTTTACCGCGGATTTTACCGCGGATTTTACCGCGG 6189
 QY 4719 CAGACAGTGCATCTCAGCTTGGATCCCACTTCCACCTTCCACCTTCCACCTTCCACCT 4778
 Db 6190 CAGACAGTGCATCTCAGCTTGGATCCCACTTCCACCTTCCACCTTCCACCTTCCACCT 6249
 QY 4779 GACGGGTGTCGGCTTCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4838
 Db 6250 GACGGGTGTCGGCTTCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6309
 QY 4839 AGGTTTGTGATCTCCAGGAGAAAGCGGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAG 4898
 Db 6310 AGGTTTGTGATCTCCAGGAGAAAGCGGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAG 6369
 QY 4899 TGTATGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4958
 Db 6370 TGTATGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6429
 QY 4959 GGGCTTACCTTAATATACACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5018
 Db 6430 GGGCTTACCTTAATATACACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6489
 QY 5019 AGCGTTCACAGGCTCACCACATATAGTCCCACTTCCCTGCTCCAGACTTAAACAGGCA 5078
 Db 6490 AGCGTTCACAGGCTCACCACATATAGTCCCACTTCCCTGCTCCAGACTTAAACAGGCA 6549
 QY 5079 GGAGACAACTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 5138
 Db 6550 GGAGACAACTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 6609
 QY 5139 CCACCTTCATCGTGGGACCAATATGGAAGTGTCTCATACCGCTGAAACCTTACCTGACAC 5198
 Db 6610 CCACCTTCATCGTGGGACCAATATGGAAGTGTCTCATACCGCTGAAACCTTACCTGACAC 6669
 QY 5199 GGGCCAAACCCCTGCTGTATAGGCTAGGAGCGCTTCCAAATATAGGCTCATCTCCACACAC 5258
 Db 6670 GGGCCAAACCCCTGCTGTATAGGCTAGGAGCGCTTCCAAATATAGGCTCATCTCCACACAC 6729
 QY 5259 CCCATTAATAATACATCATGCGATGCTGCGGTGACCTTGGAGGCTGCTGCTGCTGCTGCTG 5318

Db 6730 CCCATAACCAATATCATCATGCGATGTCATGTCGCGCTGACCTGGAGGTGCTCATCGAGCAACC 6789
Qy TGGGTGCTGTAGCGGAGTCTCTTGAGCTTTGGCCGCTACTGCTGACGACAGGCGAGT 5378
Db TGGGTGCTGTAGCGGAGTCTCTTGAGCTTTGGCCGCTACTGCTGACGACAGGCGAGC 6849
Qy GTGGTCAATTTGGCGAGGATCATCTTGTCCGGGAAGCAGCTGCTGCTTCCCGACAGGGAA 5438
Db GTGGTCAATTTGGCGAGGATCATCTTGTCCGGGAAGCAGCTGCTGCTTCCCGACAGGGAA 6909
Qy GTCTCTTACAGGAGTTCGATGAGATGGAAGAGTGTGCTTCAACAACTTCTTACATCGAG 5498
Db GTCTCTTACAGGAGTTCGATGAGATGGAAGAGTGTGCTTCAACAACTTCTTACATCGAA 6969
Qy CAGGGAATGAGCTCGCGAGCAATTCAGCAAGAGGCGCTCGGGTTGTTCGCAACGGCC 5558
Db CAGGGAATGAGCTCGCGAGCAATTCAGCAAGAGGCGCTCGGGTTGTTCGCAACGGCC 7029
Qy ACCAAGCAAGCGAGGCTGCTGCTCCGCTGCTGAGTCCAGTGGCGAGCGCTTGAACC 5618
Db ACCAAGCAAGCGAGGCTGCTGCTCCGCTGCTGAGTCCAGTGGCGAGCGCTTGAAGCC 7089
Qy TTCTGCGGGAAGACATGTGAATTTCAACGCGGAATACAGTACCTAGAGGCTTATCC 5678
Db TTCTGCGGGAAGACATGTGAATTTCAACGCGGATACAGTATTTAGCAGGCTTGTCC 7149
Qy ACTCTGCTGGAACCCCGGATAGCATCATTTGATGCGATTTACAGCTTCTATCACTAGC 5738
Db ACTCTGCTGGAACCCCGGATAGCATCATTTGATGCGATTTACAGCTTCTATCACTAGC 7209
Qy CCGCTCACCAACCCAAACCCCTCTGTTTAACTTTGCGGGGATGGTGGCTGCCCAA 5798
Db CCGCTCACCAACCCAAACCCCTCTGTTTAACTTTGCGGGGATGGTGGCGGCCAA 7269
Qy CTCCTCTCCAGGCTGCTGCTGCTTCTGCGGCGCGCATGCGCGAGCGGCTGTT 5858
Db CTCCTCTCCAGGCTGCTGCTTCTGCGGCGCGCATGCGCGAGCGGCTGTT 7329
Qy GGCAGCATAGGCTTGGGAAGTGTCTGTGACATCTTGGCGGCTATGGCGAGGGGTA 5918
Db GGCAGCATAGGCTTGGGAAGTGTCTGTGATATTTGGCAGGTTATGGAGCAGGGGTG 7389
Qy GCCCGGCACTGTGGCTTTAAGTCAAGAGCGCGAGGTGCGCTTCAACCGAGGACTG 5978
Db GCAGCGCGCTCTGTGGCTTTAAGTCAATGAGCGCGAGATGCGCTTCAACCGAGGACTG 7449
Qy GTCAACTTACTCCCTGCGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6038
Db GTTAACTTACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7509
Qy GCAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6098
Db GCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7569
Qy ATAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6158
Db ATAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7629
Qy GCTCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6218
Db GCTCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7689
Qy CTCACCAAGTGGATTAATGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6278
Db CTTCACAGTGGATTAATGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7749
Qy GTTTGGGATGGAATGCAAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6338
Db GTTTGGGATGGAATGCAAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7809
Qy CTGCGCGGTTTACCGGAGTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6398
Db CTGCGCGGTTTACCGGAGTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7899

Qy CGGGGAGCGGATCATGCAAAACCACTGCTCCCATGCGGAGCAAGATGCGCGACATGTC 6458
Db CGGGGAGCGGATCATGCAAAACCACTGCTCCCATGCGGAGCAAGATGCGCGACATGTC 7929
Qy AAAAAGCGTTTCCATGAGGATCTGAGGCTTAAAGCTGAGCAACACGTTGGGACGAAAG 6518
Db AAAAAGCGTTTCCATGAGGATCTGAGGCTTAAAGCTGAGCAACACGTTGGGACGAAAG 7989
Qy TTCCCATCAACGATACACCGGAGCTTGCACACCTCCCGCGCGCCCACTATTTC 6578
Db TTCCCATCAACGATACACCGGAGCTTGCACACCTCCCGCGCGCCCACTATTTC 8049
Qy AGGGCGCTTGGCGGCTGCTGAGGAGTACGTTGAGGCTTACGCGGTGGGGATTC 6638
Db AGGGCGCTTGGCGGCTGCTGAGGAGTACGTTGAGGCTTACGCGGTGGGGATTC 8109
Qy CATCTGAGCGGCTGACCACTGACAGCTTAAAGTGGCTGCTGCGGTTGCGGCGCC 6698
Db CATCTGAGCGGCTGACCACTGACAGCTTAAAGTGGCTGCTGCGGTTGCGGCGCC 8169
Qy GAATTTTCAAGGAGTGGATGCGGTTGACAGCTGCTGCGGCTGCTGCGGCTGCTGCGG 6758
Db GAATTTTCAAGGAGTGGATGCGGTTGACAGCTGCTGCGGCTGCTGCGGCTGCTGCGG 8229
Qy CTCTTACGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6818
Db CTCTTACGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8289
Qy CTCTTACGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6878
Db CTCTTACGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8349
Qy CATTACGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6938
Db CATTACGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8409
Qy AGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6998
Db AGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8469
Qy GACTCCCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7058
Db GACTCCCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8529
Qy AACATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7118
Db AACATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8589
Qy CACGCGAGGAGGATGAGAGGAGATATCCGTCGCGGAGATCTGCGGAAATCCAGG 7178
Db CACGCGAGGAGGATGAGAGGAGATATCCGTCGCGGAGATCTGCGGAAATCCAGG 8649
Qy AGTTTCCCTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7238
Db AGTTTCCCTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8709
Qy TCTTGAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7298
Db TCTTGAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8769
Qy AAGGCTTCCCAATACCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7358
Db AAGGCTTCCCAATACCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8829
Qy GTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7418
Db GTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8889
Qy GTTATAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7478
Db GTTATAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8949

QY 7479 TCCGAGTTGAGTCGTTACTCTCTCCATCGCCCCCTTTGAAGGCGAGCCGGGGACCCCGGAT 7538
Db 8950 TCCGAGTTGAGTCGTTACTCTCTCCATCGCCCCCTTTGAAGGCGAGCCGGGGATCCCGAT 9009
QY 7539 CTCAGGAGGGGTCCTTGGTCTACCGTGTAGTGGAGGCTAGTGAGGATGTGCTGCTGC 7598
Db 9010 CTCAGGAGGGGTCCTTGGTCTACCGTGTAGTGGAGGCTAGTGAGGACCTGCTGCTGCTGC 9069
QY 7599 TCAATGTCTATACGTGGACAGCGCCCTGATCAGCCCATGCGCTCGCGAGGAGAAAGTAA 7658
Db 9070 TCGATGTCTATCACATGAGACAGCGCCCTGATCAGCCCATGCGCTCGCGAGGAAACCAAG 9129
QY 7659 CTGCCCCATCAACCGTTGAGCAACTCTTTGCTGCGTCAACCAACAATGTCTACGCCACA 7718
Db 9130 CTGCCCCATCAATGCACTGAGCAACTCTTTGCTGCGTCAACCAACAATGTCTATGCTACA 9189
QY 7719 ACATCCGCGAGCGCAAGCCTCCGGCAGAGAGAGGTCACTTTTGACAGATTGCAAGTCTTG 7778
Db 9190 ACATCTCGCAGCGCAAGCCTCGGGCAGAGAGAGGTCACTTTTGACAGATTGCAAGTCTTG 7778
QY 7779 GATGATCATTTACGGGACGTACTCAAGGAGATGAAGGCGAAGCGTCCACAGTTAAGGCT 7838
Db 9250 GACGACCACTACCGGACGTGTCAAGGAGATGAAGGCGAAGCGTCCACAGTTAAGGCT 9309
QY 7839 AGCTTCTATCTATAGAGAGGCTGCAAGCTGAGCCGCCCACTTTGGGCCAAATCCAAA 7898
Db 9310 AAACCTTCTATCCGTGAGGAGCCTGTAGCTGAGCCGCCCACTTTGGGCCAAGATCTAAA 9369
QY 7899 TTTGGCTATGGGGCAAGGACGTCGGGACCTATCCAGCAGGCGCGTTAACACATCCGC 7958
Db 9370 TTTGGCTATGGGGCAAGGACGTCGGGACCTATCCAGCAGGCGCGTTAACACATCCGC 9429
QY 7959 TCCGTGTGGGAGACTTGTGTGAAGACACTGTAACCAACCAATTTGACACCACTCATGCGCA 8018
Db 9430 TCCGTGTGGGAGACTTGTGTGAAGACACTGTAACCAACCAATTTGACACCACTCATGCGCA 9489
QY 8019 AAAAGTGAAGTTTCTGGCTCCAAACAGAGAGGAGCGCCGAAGCAGCTGCGCTTATC 8078
Db 9490 AAAAATGAGGTTTCTGGCTCCAAACAGAGAGGAGCGCGCAAGCAGCTGCGCTTATC 9549
QY 8079 GTATTCCAGACCTGGGAGTCTGTATGCGAGAGATGGCCCTTTAGACGTGCTTCC 8138
Db 9550 GTATTCCAGATTTGGGGTCTGTGTGCGAGAAATGGCCCTTTAGCATGTGCTTCC 9609
QY 8139 ACCCTTCTCAGGCGCTGATGCGCTCTCATACGATTTCAATATCTCCCAAGCAGCGG 8198
Db 9610 ACCCTTCTCAGGCGCTGATGCGCTCTCATACGATTTCAATATCTCTCTGGACAGCGG 9669
QY 8199 GTGAGTTCTGTGTAATACCTGGAATCAAGAAATGCCCTATGCGGTTCTCATATGAC 8258
Db 9670 GTGAGTTCTGTGTAATGCTGGAATGCTGGAAGCGAAGAAATGCCCTATGCGGTTCTCATATGAC 9729
QY 8259 ACCGCTCTTTTGACTCAACGCTCACTGAGAGTGACATTCGTTGTGAGAGTCAATTTAC 8318
Db 9730 ACCGCTCTTTTGACTCAACGCTCACTGAGATGACATTCGTTGTGAGAGTCAATTTAC 9789
QY 8319 CAATGTGTGACTGTGGCCCCGAGCCAGACAGGCCATAAGGTGCTCACAAGAGCGGCTT 8378
Db 9790 CAATGTGTGACTGTGGCCCCGAGCCAGACAGGCCATAAGGTGCTCACAAGAGCGGCTT 9849
QY 8379 TACATCGGGGTCCTGACTCACTCAAAAGGCGAGAACTGCGGTTATCGCGGTGCGC 8438
Db 9850 TACATCGGGGTCCTGACTCACTCAAAAGGCGAGAACTGCGGTCATCGCGGTGCGC 9909
QY 8439 GCAAGTGGCGTCTGACGACTAGCTGGGTAATACCTCTCATCTTACTTTGAAGGCCACT 8498
Db 9910 GCGAGCGGTGTACTGACGACGAGCTGCGGTAATACCTCTCATCTTACTTTGAAGGCCGCT 9969
QY 8499 GCAGCCTGTGAGTGTGCAAGCTCCAGGACTGCAAGTGTGCTGAAACGAGAGACGCTT 8558
Db 9970 GCGGCTGTGAGTGTGCAAGCTCCAGGACTGCAAGTGTGCTGAAACGAGAGACGCTT 10029
QY 8559 GTCCTTATCTGTGAAGCGGCGGAACCCAGGAGGATGCGGCGGCTTACGAGCCTTCAAG 8618

RESULT 10

Db 10030 GTGCTTATCTGTGAAGCGCGGGGACCCAAAGAGGACGAGCGGCTACGGGCTTTCAG 10089
QY 8619 GAGGCTATGACTAGTATTTCCGCCCGCCCGGGGATCCGCCCAACCAAGAAATACGACCTG 8678
Db 10090 GAGGCTATGACTAGTATTTCCGCCCGCCCGGGGACCCGCCCAACCAAGAAATACGACCTG 10149
QY 8679 GAGGCTATGACTAGTATTTCCGCCCGCCCGGGGATCCGCCCAACCAAGAAATACGACCTG 8738
Db 10150 GAGTGTGATAACATCATGCTCTCTCAATGTGTCACTCGCGCAGCATCTCTGCGCAAAAG 10209
QY 8739 GTATATCTACTCTCACCGGTGACCCCAACCCCTTTGACCGGGTGTGCGGAGACAGCT 8798
Db 10210 GTGTACTATCTCACCGGTGACCCCAACCCCTTTGCGGGGCTGCGGTGGAGACAGCT 10269
QY 8799 AGACACATCCAAATCAATCTCTTGGGTAGGCAATATCATGTATGTATGCGCCCACTTGG 8858
Db 10270 AGACACATCCAAATCAATCTCTTGGGTAGGCAATATCATGTATGTATGCGCCCACTTGG 10329
QY 8859 GCAAGGATGATCTGATGACTCACTTTTCTCCATCTCTTAGCTCAAGAGCAATTTGAA 8918
Db 10330 GCAAGGATGATCTGATGACTCACTTTTCTCCATCTCTTAGCTCAAGAGCAATTTGAA 10389
QY 8919 AAAGCCCTGGATTTGATGATCTAGCGGCTTGTACTCCATTTAGCCCACTTTGACCTACT 8978
Db 10390 AAAGCCCTTAGATTTGATGATCTAGCGGCTTGTACTCCATTTAGCCCACTTTGACCTACT 10449
QY 8979 CAGATCATTTGAAGACTCCATGCTTTAGCGCAATTTACACTCCACAGTTACTTCCAGT 9038
Db 10450 CAGATCATTTGAAGACTCCATGCTTTAGCGCAATTTTCACTCCATGTTACTTCTCCAGT 10509
QY 9039 GAGATCAATAGGTTGGCTTCATGCTCAGGAACTTTGGGTTACCACTTGGGAACTTGG 9098
Db 10510 GAGATCAATAGGTTGGCTTCATGCTCAGGAACTTTGGGTTACCACTTGGGAACTTGG 10569
QY 9099 AGACATCGGGCAGAAAGTGTCCGGCTTAAGTACTGTCTCCAGGGGGGAGGCGGCCACT 9158
Db 10570 AGACATCGGGCAGAAAGTGTCCGGCTTAAGTACTGTCTCCAGGGGGGAGGCGGCCACT 10629
QY 9159 TGTGGCAGATACCTCTTTAACTGGGCGAGTAAGGCAAAAGCTTAAAGCTTAACTTCCATCCG 9218
Db 10630 TGTGGCAGATACCTCTTTAACTGGGCGAGTAAGGCAAAAGCTTAAAGCTTAACTTCCATCCG 10689
QY 9219 GCCCGTCCAGCGCTGCTGTGCTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9278
Db 10690 GCT 10749
QY 9279 TATCAGACCTGTCTGCTGCGGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9338
Db 10750 TATCAGACCTGTCTGCTGCGGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10809
QY 9339 GTAGGGGTAGGCAATTTACCTGCTCCCAACCGATGAACGGGAGCTAACCACTCCAGGCC 9398
Db 10810 GTAGGGGTAGGCAATTTACCTGCTCCCAACCGATGAACGGGAGCTAACCACTCCAGGCC 10869
QY 9399 TTAAGCC-----ATTTCCTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 9448
Db 10870 AATAGGCAATCTGTTTTTTTCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 10929
QY 9449 TTTTCTTCT 9508
Db 10930 TTTTCTTCT 10989
QY 9509 CTTAGCCCTTAGTCAAGGCTAGCTGTGAAGGTCGCTGAGCCGATCACTGCAAGAGTGC 9568
Db 10990 CTTAGCCCTTAGTCAAGGCTAGCTGTGAAGGTCGCTGAGCCGCTTCACTGACAGAGTGC 11049
QY 9569 TGATACTGCGCTCTCTGACAGATCATG 9595
Db 11050 TGATACTGCGCTCTCTGACAGATCAAGT 11076

```

US-09-539-601-19
; Sequence 19, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartschlagher, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; EARLIER FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 11076
; TYPE: DNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(341)
; OTHER INFORMATION: construct I389/Core-3',/9-13F
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342)..(1193)
; OTHER INFORMATION: hepatitis C virus core - neomycin
; OTHER INFORMATION: phosphotransferase fusion protein
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (1202)..(1812)
; OTHER INFORMATION: internal ribosome entry site from
; OTHER INFORMATION: encephalomyocarditis virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1813)..(10845)
; OTHER INFORMATION: hepatitis C virus polyprotein from core to
; OTHER INFORMATION: nonstructural protein NS5B; carries cell
; OTHER INFORMATION: culture-adaptive mutations from clone 9-13F
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (10846)..(11076)
US-09-539-601-19

Query March
Best Local Similarity 82.3%; Score 7899.4; DB 4; Length 11076;
Matches 842; Conservative 0; Mismatches 836; Indels 10; Gaps 1;

QY 339 ACCATGAGCGACGAATCCTAAACCTCAAGAGAAAACCAAAAGCTATACACCAACCGCGGCCA 398
Db 1810 ACCATGGCGACGATCTCTAAACCTCAAGAGAAAACCAAAAGCTATACACCAACCGCGGCCA 1869

QY 399 CAGGACGTCGAATTCCTCCGGCGGCTGGTCAGATCGTTGGTGGAGTTTACCTGTCGCGGC 458
Db 1870 CAGGACGTCGAATTCCTCCGGCGGCTGGTCAGATCGTTGGTGGAGTTTACCTGTCGCGGC 1929

QY 459 AGGGGGCCCCAGGTGGGTGGTCGGCGGCACCTAGGAAGCTTCCGAGCGGTGCGAACCTCGT 518
Db 1930 AGGGGGCCCCAGGTGGGTGGTCGGCGGCACCTAGGAAGCTTCCGAGCGGTGCGAACCTCGT 1989

QY 519 GGAAGGCGACCAACCTATCCCAAGAGCTCGCGACCCGAGGCGAGGCGCTGGGCTCAGCCC 578
Db 1990 GGAAGGCGACCAACCTATCCCAAGAGCTCGCGACCCGAGGCTGGGCGCTCAGCCC 2049

QY 579 GGGTACCCCTGGGCCCTCTATGGCAATGAGGCCCTGGGGCTGGGCGAGATGGCTCTGTCA 638
Db 2050 GGGTACCCCTGGGCCCTCTATGGCAATGAGGCCCTGGGGCTGGGCGAGATGGCTCTGTCA 2109

QY 639 CCCCGCGGCTCCGGGCTAGTTGGGGCCCCACGAGACCCCGCGGTAGGTGCGGTAACTTG 698
Db 2110 CCCCGTGGCTCTCGGCCCTAGTTGGGGCCCCACGAGACCCCGCGGTAGGTGCGGTAACTTG 2169

QY 699 GGTAAGGTATCGATACCCCTACATCGGGCTTCGCGATCTCATGGGGTACATTCGCTC 758
Db 2170 GGTAAGGTATCGATACCCCTACATCGGGCTTCGCGATCTCATGGGGTACATTCGCTC 2229

```

QY 1839 TCGCAGCTGTGTGTCAGTGTATTTTACACCCCAAGCCCTGTGTGTGGGACACCC 1898
Db 3310 GCGCAGGTGTGTGTCAGTGTACTGCTTACCCCAAGCCCTGTGTGTGGGACACCC 3369
QY 1899 GATCCTTCCGCTGTCCCTACGTATAGCTGTGGGGGAGAAATGACACAGAGCTGATGCTCTC 1958
Db 3370 GACCGGTTCCGGGTCTTACGTACAGTTGGGGGAGAAATGACACAGAGCTGATGCTCTT 3429
QY 1959 AACAAACAGCGCTCCGCCCAAGGCAACTGGTTTGGCTGTATGATGAATAGTACTGGG 2018
Db 3430 AACAAACAGCGCGCCCGCAGGCAACTGGTTTGGCTGTATGATGAATAGTACTGGG 3489
QY 2019 TTCACTAAGAGCTGGGAGGTCCCGGTGTACATCGGGGGGTCTGTTAAACCGACCTTG 2078
Db 3490 TTCACCAAGAGCTGGGGGGGGGGGGGTGTAACATCGGGGGGATCGGCAATAAAACCTTG 3549
QY 2079 ATCTCCGCCACGAGCTGTCTCGGAAGACACCCGAGGCTTACTTACACAAATATGTGGCTCG 2138
Db 3550 ACCTGCCCAACGAGCTGTCTCGGAAGACACCCGAGGCACTTACACCAAGTGTGGTTGG 3609
QY 2139 GGGCCCTGGTTGACACCTAGGTGCTTAGACTACCATACAGCTTTGGCACTTACCCC 2198
Db 3610 GGGCCCTGGTTGACACCCAGATGCTTGGTCCACTACCCATACAGGCTTTGGCACTTACCCC 3669
QY 2199 TGCACCTCTCAATTTTCCATCTTTAAGCTTAGGATGTATCTGGGGGGCTGGAGACACAGG 2258
Db 3670 TGCACCTCTCAACTTTTACATCTTTAAGGTTAGGATGTATCTGGGGGGAGTGGAGACACAGG 3729
QY 2259 CTCGAATGCGCATGCAATTTGACCTGAGAGAGCGCTGTAACTTGGAGACAGGAGATAGG 2318
Db 3730 CTCGAATGCGCATGCAATTTGACCTGAGAGAGCGTGTAACTTGGAGACAGGAGATAGG 3789
QY 2319 TCAGAACTCAGCCGCTGCTGTCTACACAGAGTGGGAGATCTCCCTGTGCTTTC 2378
Db 3790 TCAGAGCTTAGCCCGCTGCTGTCTACACAGGAGTGGCAGGTATTGCCCTGTGCTTTC 3849
QY 2379 ACCACCTTACCGGCTTTATCCACTGTTTGTATCCATCTCCATCAGAACTGTGAGAGTG 2438
Db 3850 ACCACCTTACCGGCTTCTGTCCACTGGTTGTATCCATCTCCATCAGAACTGTGAGAGTG 3909
QY 2439 CAATACCHTACGCTGAGGCTGAGGCTGTCTCTTTGCAATCAAAATGGGAGTACATC 2498
Db 3910 CAATACCTGTACGCTATAGGCTCGCGGCTGTCTCTTTGCAATCAAAATGGGAGTATGTC 3969
QY 2499 CTGTGCTTTTCTCTCTGCGAGACGCGGCTGTGCTCTCTGATGATGATGCTG 2558
Db 3970 CTGTGCTTTCTCTCTGCGGAGACGCGGCGCTGTGCTCTCTGATGATGATGCTG 4029
QY 2559 CTGATAGCCACGAGCTGAGCGCGCTTACAGAACTTGTGTGCTCTCAATGCGGCTGCTG 2618
Db 4030 CTGATAGCTCAGCTGAGCGCGCTTACAGAACTTGTGTGCTCTCAATGCGGCTGCTG 4089
QY 2619 GCGGAGCGGATGATCTCTCTCTTCTGTGCTCTCTGCGGCTGCTGATCAATTAAG 2678
Db 4090 GCGGAGCGGATGATGATCTCTCTCTTCTGTGCTCTCTGCGGCTGCTGATCAATTAAG 4149
QY 2679 GCGAGGCTGCTCTGCGGCGGCTGATGCTTTTATGCGGCTGATGCGGCTGCTGCTG 2738
Db 4150 GCGAGGCTGCTCTGCGGCGGCTGATGCTTTTATGCGGCTGATGCGGCTGCTGCTG 4209
QY 2739 CTACTGCGGTTACCAACACGAGCTTACGCTTGGACCGGAGAGTGGCTGCTGCGGG 2798
Db 4210 CTGCTGGGTTTACCAACACGAGCTTACGCTTGGACCGGAGAGTGGCTGCTGCGGG 4269
QY 2799 GGTGCGGTTCTGTAGGCTGCTGCTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2858
Db 4270 GCGCGGTTTCTGTAGGCTGCTGCTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 4329
QY 2859 ACTAGGCTCATATGGTGTACATATCTTATACCGAGCGGAGGCGACATGCAAGTG 2918
Db 4330 GCTAGGCTCATATGGTGTACATATCTTATACCGAGCGGAGGCGACATGCAAGTG 4389
QY 2919 TGGCTCCCCCTTCAAGCTTCCGGGAGCGCGCATGCCATCATCTCTCTCACCTGTGCG 2978

Db 4390 TGGATCCCCCTTCAACGTTCCGGGGGCGCGCATGCCGCTCATCTCTCACTGCGCG 4449
QY 2979 GTTCATCCAGAGTTAAATTTTTCATCACCAAACTCTGCTCGCATCTGCGGCGGCTC 3038
Db 4450 ATCCACCCAGAGCTAAATCTTTTACCATCACCAAACTCTGCTCGCATCTGCGGCTC 4509
QY 3039 ATGTGTCTCCAGGCTGGCAATACGAGAGTGCCTTACTTCTGTCGCGCTCAAGGGCTCAT 3098
Db 4510 ATGTGTCTCCAGGCTGGTATTAACCAAAAGTGCCTTACTTCTGTCGCGCACAGGGCTCAT 4569
QY 3099 GTCGATGCTATGTATGTGCGAAAGTCCCGGGGCTCATTTATGTCCAAATGGTCTTCATG 3158
Db 4570 GTCGATGCTATGTATGTGCGAAAGTCCCGGGGCTCATTTATGTCCAAATGGTCTTCATG 4629
QY 3159 AAGCTGGGCGCGTACAGGTACGTAGTGTATTAACCATCTTACCCGCTGCGGAGCTGG 3218
Db 4630 AAGTTGGCGGCTACGTAGGCTACGTATTAACCATCTTACCCGCTGCGGAGCTGG 4689
QY 3219 GCCCACGCGGCTTACGAGACCTTTCGCTGGCGGTAGAGCCGCTCTTCTCCGCGCATG 3278
Db 4690 GCCCACGCGGCTTACGAGACCTTTCGCTGGCGGTAGAGCCGCTCTTCTCTGATGATG 4749
QY 3279 GAGACCAAGTCTATCACTGCGGAGCAGACACCGCTGGTGTGGGACATCATCTTGGGT 3338
Db 4750 GAGACCAAGTCTATCACTGCGGAGCAGACACCGCGGCTGTGGGACATCATCTTGGGC 4809
QY 3339 CTACCCGCTTCCGCGCGAAAGGGGAGAGATATTTTGGGACCGCTGATAGTCTCGAA 3398
Db 4810 CTGCGGCTTCCGCGCGGAGGAGATATCTTGGGACCGGAGACAGCTTGA 4869
QY 3399 GGGCAAGGCTGGGACCTTCTTGGCGCCCATCAAGGCTTACTCCCAACAAAGCGGGGCGTA 3458
Db 4870 GGGCAGGCTGGGACCTTCTTGGCGCCCATCAAGGCTTACTCCCAACAAAGCGGAGCTTA 4929
QY 3459 CTGCTGTGCTATCATCACTAGCTTCAAGGCGCGGAGCAAGAACAGTGTGAGGAGGCT 3518
Db 4930 CTGCTGTGCTATCATCACTAGCTTCAAGGCGCGGAGCAAGAACAGTGTGAGGAGGCT 4989
QY 3519 CAAGTGTGCTTACCGCAACACATCTTCTGCGGACCTGATCAAGCGGCTGCTGCTG 3578
Db 4990 CAAGTGTGCTTACCGCAACACATCTTCTGCGGACCTGATCAAGCGGCTGCTGCTG 5049
QY 3579 ACTGTCTACCATGCGCTGTGCTGAGACCTTACCGGCTTCCAAAGGTCCAAATCACCCAA 3638
Db 5050 ACTGTCTATCATGCTGCGGCTCAAGACCTTTCGCGGCTTCCAAAGGCTTCCAAATCACCCAA 5109
QY 3639 ATGTACACCATGTAGACCTTGGACCTTGTGCTGCGGCTGCGGCGGCTGCTGCTG 3698
Db 5110 ATGTACACCATGTAGACCTTGGACCTTGTGCTGCGGCTGCGGCGGCTGCTGCTG 5169
QY 3699 ATGACACCATGTAGACCTTGGACCTTGTGCTGCGGCTTGTGCTGCGGCTGCTGCTG 3758
Db 5170 TTGACACCATGTAGACCTTGGACCTTGTGCTGCGGCTTGTGCTGCGGCTGCTGCTG 5229
QY 3759 ATTCCGCTGCGCGGAGCGGACAGCAGGAGGAGTCTACTCTCCCGGAGGCTGCTGCTG 3818
Db 5230 ATTCCGCTGCGCGGAGCGGACAGCAGGAGGAGTCTACTCTCCCGGAGGCTGCTGCTG 5289
QY 3819 TACTGAAAGCTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3878
Db 5290 TACTGAAAGCTTCTGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5349
QY 3879 TTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3938
Db 5350 TTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5409
QY 3939 TCTATGAAACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3998
Db 5410 TCTATGAAACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5469
QY 3999 CCGGAGACATTCGAATGTGGGACATCTGACGCTTCTTCTGCGGAGCGGCAAGACCAAA 4058

Dd 5470 CCGCAGACATTCAGGTGGCCCACTCTACACGCGCCCTACTGTAGCGGCAAGAGCACTAAG 5529
Qy 4959 GTGCCGGCTGCGTATGACAGCCCAAGGGTACAAAGTGTCTGCTCTGAAACCCCTCGCTGGCC 4118
Dd 5530 GTGCCGGCTGCGTATGACAGCCCAAGGGTATAAGGTGCTTGCTTGAACCCCTCGCTGGCC 5589
Qy 4119 GCGACCTTAGGGTTTGGGGCGTATATGTCCAAAGGCACAGGTATGACCCCTAAACATGAGA 4178
Dd 5590 GCGACCCCTTAGGTTCGGGGCGTATATGTCTAAGGCACATGATATCGACCCCTAAACATCAGA 5649
Qy 4179 ACTGGGTAAAGACCAATTAACACGGGGCGCTCCATTACCTACTCCACCTATGCAAGTTC 4238
Dd 5650 ACCGGGTAGGACCAATACACAGGGTCCGCCCATCAGCTACTCCACCTATGCGCAAGTTT 5709
Qy 4239 CTTGCCGACGCTGCTGTTCTGGGGCGCCCTPATGACATCAATAATATGATGAGTGGCCAC 4238
Dd 5710 CTTGCCGACGCTGCTGTTCTGGGGCGCCCTPATGACATCAATAATATGATGAGTGGCCAC 5769
Qy 4299 TCACCTGACTCGACTACCATCTTGGGATCGGCACAGTCTCTGGACCAAGCGGAGCGCT 4358
Dd 5770 TCACCTGACTCGACCATCATCTTGGGATCGGCACAGTCTCTGGACCAAGCGGAGCGCT 5829
Qy 4359 GGAGCGGCTGCTGCTCGCCACCGCTACACTCCGGGATCGGTTACCGTGCACAC 4418
Dd 5830 GGAGCGGCTGCTGCTCGCCACCGCTACCGCTCCGGGATCGGTTACCGTGCACAC 5889
Qy 4419 CCCAATATCGGAAATATGGCTGTCTCAAAATAGAGATCCCTTCTATGGCAAGCC 4478
Dd 5890 CCAAAATATCGAGGAGTGGCTCTCTCCAGCACTGGAGAAATCCCTTCTATGGCAAGCC 5949
Qy 4479 ATCCCATATGAGGCATCAAGGGGGGAGGATCTCATTTCTCCGATTCGCAATCCCAAGAGAA 4538
Dd 5950 ATCCCATATGAGGCATCAAGGGGGGAGGATCTCATTTCTCCGATTCGCAATCCCAAGAGAA 6009
Qy 4539 TGTGACGAGCTCGCCGAAAGCTGACAGGCTCGGACTGGAAGCTGTAGCATATTACCGG 4598
Dd 6010 TGTGATGAGCTCGCCGAAAGCTATCGGCTCGGACTCAATGCTGTAGCATATTACCGG 6069
Qy 4599 GSCCTTATGTCGCTCATACCGCTATACCGCTATCGGAGAGTGTGTCGGCAACAGAGCT 4658
Dd 6070 GSCCTTATGTCGCTCATACCGCTATACCGCTATCGGAGAGTGTGTCGGCAACAGAGCT 6129
Qy 4659 CTAATGACGGGTTTACCGGCTATTTGACTCATGATGAGTGTGCAATCATATGTCAC 4718
Dd 6130 CTAATGACGGGTTTACCGGCTATTTGACTCATGATGAGTGTGCAATCATATGTCAC 6189
Qy 4719 CAGACATGCACTTACGCTTGGATCCACCTTCACTTCACTTCACTTCACTTCACTTCACT 4778
Dd 6190 CAGACATGCACTTACGCTTGGATCCACCTTCACTTCACTTCACTTCACTTCACTTCACT 6249
Qy 4779 GACCGGTGTCGGCTCGCAACGCGAGTGAATCTGGCGGGGTAGAGTGGATCTAC 4838
Dd 6250 GACCGGTGTCGGCTCGCAACGCGAGTGAATCTGGCGGGGTAGAGTGGATCTAC 6309
Qy 4839 AGTTTGTGACTCCAGGAGAACGGCCCTCGGCAATGTCGATCTTCTGCTGCTGTCGAG 4898
Dd 6310 AGTTTGTGACTCCAGGAGAACGGCCCTCGGCAATGTCGATCTTCTGCTGCTGTCGAG 6369
Qy 4899 TGCTATGACCGGGCTGTGCTTGTATGATCAACCGCCCTGAGACCTCGGTTAGGTTG 4958
Dd 6370 TGCTATGACCGGGCTGTGCTTGTATGATCAACCGCCCTGAGACCTCGGTTAGGTTG 6429
Qy 4959 CGGCTTACCTAAATACACAGAGGTTTCCCGTCTGCGAGGAGTGAATCTGGAGTTCTGGAG 5018
Dd 6430 CGGCTTACCTAAATACACAGAGGTTTCCCGTCTGCGAGGAGTGAATCTGGAGTTCTGGAG 6489
Qy 5019 AGCGTTTACAGGCTTCAACCATAGATGCGCACTTCTGTCGCAAGTGAATCAACAGGCA 5078
Dd 6490 AGCGTTTACAGGCTTCAACCATAGATGCGCACTTCTGTCGCAAGTGAATCAACAGGCA 6549
Qy 5079 GGAGACAACTTCTTACCTGTGGGCAATCAAGCTACAGTGTGCGCCAGGCTCAAGCT 5138
Dd 6550 GGAGACAACTTCTTACCTGTGGGCAATCAAGCTACAGTGTGCGCCAGGCTCAAGCT 6609

Qy 5139 CCACCTCCATCGTGGGACCAAAATGTGGAACTGTCTCATACGGCTGAAACCTCACATGCAC 5198
Dd 6610 CCACCTCCATCGTGGGACCAAAATGTGGAACTGTCTCATACGGCTGAAACCTCACATGCAC 6669
Qy 5199 GGGCCAAACCCCTGCTGTATAGGCTAGGAGCCGTCCAAAATAGGCTCACTCAGACAC 5258
Dd 6670 GGGCCAAACCCCTGCTGTATAGGCTAGGAGCCGTCCAAAATAGGCTCACTCAGACAC 6729
Qy 5259 CCACCTAAATATACATCATGCGATCAGTGTGGCTGACCTGAGTGTGCTCACTAGACAC 5318
Dd 6730 CCACCTAAATATACATCATGCGATCAGTGTGGCTGACCTGAGTGTGCTCACTAGACAC 6789
Qy 5319 TGGGCTGCTGTAGGCTAGGCTGCTTGGCGCATCTTGGCGCATCTTGGCGCATCTTGGCGCAT 5378
Dd 6790 TGGGCTGCTGTAGGCTAGGCTGCTTGGCGCATCTTGGCGCATCTTGGCGCATCTTGGCGCAT 6849
Qy 5379 GTGGTATGCTGGGCAAGGATCATCTTGTGCGGAAAGCGCTGCTGCTGCGGAAAGCGGAA 5438
Dd 6850 GTGGTATGCTGGGCAAGGATCATCTTGTGCGGAAAGCGGCTGCTGCTGCGGAAAGCGGAA 6909
Qy 5439 GTGCTTACAGGAGTGTGATGAGTGTGAGTGTGCTTCAAACTTGTGCTTCACTAGAC 5498
Dd 6910 GTGCTTACAGGAGTGTGATGAGTGTGAGTGTGCTTCAAACTTGTGCTTCACTAGAC 6969
Qy 5499 CAGGGAATGAGCTCGCCGCAAGCAATTAAGCAAAAGCGCTGCTGCTTGTGCAAAAGCGC 5558
Dd 6970 CAGGGAATGAGCTCGCCGCAAGCAATTAAGCAAAAGCGCTGCTGCTTGTGCAAAAGCGC 7029
Qy 5559 ACCAAGCAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5618
Dd 7030 ACCAAGCAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7089
Qy 5619 TTCTGCGGCAAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5678
Dd 7090 TTCTGCGGCAAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7149
Qy 5679 ACTTCTGCTGGAACCCCGCATAGCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5738
Dd 7150 ACTTCTGCTGGAACCCCGCATAGCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7209
Qy 5739 CGCTCACCACCAAAACCCCGCATAGCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5798
Dd 7210 CGCTCACCACCAAAACCCCGCATAGCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7269
Qy 5799 CTGCTCTCTCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5858
Dd 7270 CTGCTCTCTCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7329
Qy 5859 GGCAGATGAGGCTTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5918
Dd 7330 GGCAGATGAGGCTTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7389
Qy 5919 GCGCGGCGCTGCTGCTGCTTAAAGTCAAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 5978
Dd 7390 GCGAGGCGCTGCTGCTGCTTAAAGTCAAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 7449
Qy 5979 GTCAACTTACTCCCTGCCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6038
Dd 7450 GTTAACTTACTCCCTGCCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7509
Qy 6039 GCHATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6098
Dd 7510 GCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7569
Qy 6099 ATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6158
Dd 7570 ATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7629
Qy 6159 GCTGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6218
Dd 7630 GCTGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7689

QY 6219 CTCCACAGTGGATTAAGAGGACTGCTCTACCCCATGCTCCGGCTCGTGGCTAAGGGAT 6278
Db 7690 CTTACACAGTGGATCAACAGGAGCTGCTCCACGCCATGCTCCGGCTCGTGGCTAAGAGAT 7749
QY 6279 GTTTGGATTGGATATGACAGGTTGACTGACTTCAAGACCTGGCTCCAGTCCAAACTC 6338
Db 7750 GTTTGGATTGGATATGACAGGTTGACTGACTTCAAGACCTGGCTCCAGTCCAAACTC 7809
QY 6339 CTGCCCGGTACCGGAGTCCCTTCTCTCATGCTCAACCGGGTACAGGGAGTCTGG 6398
Db 7810 CTGCCCGGTACCGGAGTCCCTTCTCTCATGCTCAACGGTACAGGGAGTCTGG 7869
QY 6399 CGGGGAGCGGCATCATGAACACCTGCTCCATGCGGAGACAGANATGCGGACATGTC 6458
Db 7870 CGGGGAGCGGCATCATGAACACCTGCTCCATGCGGAGACAGANATGCGGACATGTC 7929
QY 6459 AAAAAACGGTTCCATGAGGATCGTAGGGCTTAGAACCTGACAGCAACACGTTGGCAACG 6518
Db 7930 AAAAACGGTTCCATGAGGATCGTAGGGCTTAGAACCTGAGACCTGATGAACACGTTGGCAACG 7989
QY 6519 TTCCCATCAACGCATACACCGGACCTTTCACACCTCCCGCGGCCCCAACTATTC 6578
Db 7990 TTCCCATCAACCGGTACACCGGACCTTTCACACCTCCCGCGGCCCCAACTATTC 8049
QY 6579 AGGGCGCTATGGCGGTGCTGCTGAGGATGAGTGGAGTTACGGGTGTGGGGATTTTC 6638
Db 8050 AGGGCGCTATGGCGGTGCTGCTGAGGATGAGTGGAGTTACGGGTGTGGGGATTTTC 8109
QY 6639 CACTACGTGACGGGATGACCTGACAACTGATAAGTGCCTGACAGTTCCGAGTTCCGGCCCC 6698
Db 8110 CACTACGTGACGGGATGACCTGACAACTGATAAGTGCCTGAGTTCCGGCCCC 8169
QY 6699 GAATCTTCAACGGAGTGAAGTGGCTGACAGTACGCTCCGGCGTGCACAACT 6758
Db 8170 GAATCTTCAACGAAGTGAAGTGGCTGACAGTACGCTCCAGGTGCACAACT 8229
QY 6759 CTTCTACGGGAGAGTCAAGTTCAGTTCGGGTCAACCAATATCTTGGTGGGTGGAG 6818
Db 8230 CTTCTACGGGAGAGTCAAGTTCAGTTCGGGTCAACCAATATCTTGGTGGGTGGAG 8289
QY 6819 CTTCCATGCGAGCCGAAACCGAGGTACAGTGCCTTACTTCCATGCTCACCGATCCCTCC 6878
Db 8290 CTTCCATGCGAGCCGAAACCGAGGTACAGTGCCTTACTTCCATGCTCACCGATCCCTCC 8349
QY 6879 CACATTAACAGAGAACCGCTAAGCTAGGTGGCTAGAGGTCTCCCGCTCTTTAGCC 6938
Db 8350 CACATTAACCGGACAGCGCTAAGCTAGGTGGCTAGAGGTCTCCCGCTCTTTAGCC 8409
QY 6939 AGCTCATCAGCTAGCCAGTTCCTGGCTTCTGGCTTCTTTGAGGAGACATGCACTACCACT 6998
Db 8410 AGCTCATCAGCTAGCCAGTTCCTGGCTTCTGGCTTCTTTGAGTAAATTTTGGACTCTTTGAGCCGCTC 8469
QY 6999 GACTCCCGGAGCGTGAACCTCATCGAGGCAACCTCTTGTGGCGGACAGAGATGGCGGA 7058
Db 8470 GACTCCCGGAGCGTGAACCTCATCGAGGCAACCTCTTGTGGCGGACAGAGATGGCGGG 8529
QY 7059 AACATCACTCGGTGGATGAGAGATTAAGTGTGATTTCTTGGACTCTTTGGAACCGGTT 7118
Db 8530 AACATCACTCGGTGGATGAGAGATTAAGTGTGATTTCTTGGACTCTTTGAGCCGCTC 8589
QY 7119 CACGGGAGGGGATGAGAGGAGATTCCTGCTGGCGGAGATCCTGCGAATACTCAGG 7178
Db 8590 CACGGGAGGGATGAGAGGAGATTCCTGCTGGCGGAGATCCTGCGGAGATCCTGCGAGG 8649
QY 7179 AAGTTCCCTCAGCGTTGCCATATGGGCACGCCCGGACTCAATCTCCCTCAGCTGATAG 7238
Db 8650 AATTTCCCTCAGCGATGCCATATGGGCACGCCCGGATTAACACCTCCCTCAGCTGATAG 8709
QY 7239 TCCTGGAAGGACCGGACTACGTCTCCTCGGTGTGATGAGTGCCTATGCGCACTACC 7298
Db 8710 TCCTGGAAGGACCGGACTACGTCTCCTCGGTGTGATGAGTGCCTATGCGCACTACC 8769
QY 7299 AAGGCTCTCCCAATACCACTCCACGGAGAAAGAGGAGCGGTGTCTGCTGACAGAAATCCAA 7358

Db 8770 AAGGCCCTCCGATACCACCTCCACGGAGGAGAGACGGTTCTCTGTCAGATCTACC 8829
QY 7359 GTGTCTTCTGCTTGGCGAGCTCGCACTAAGACTTTCGGTAGCTCCGGATCGTGGCC 7418
Db 8830 GTGTCTTCTGCTTGGCGAGCTCGCACTAAGACTTTCGGTAGCTCCGGATCGTGGCC 8889
QY 7419 GTTGTAGCGGACACGGGACCGCTTCTGAGACCTGGCTCCGACGAGCGTGACAAAG 7478
Db 8890 GTTCGACGGGACACGGGACCGCTTCTGAGCAGCCCTCCGACGAGCGGACCGGGA 8949
QY 7479 TCCGACGTGAGTCTGACTCTCTCCATGCCCCCTTTGAAGGGGAGCCCGGGGACCCCAT 7538
Db 8950 TCCGACGTGAGTCTGACTCTCTCCATGCCCCCTTTGAAGGGGAGCCCGGGGATCCCAT 9009
QY 7539 CTCAGGAGCGGTCTTGTCTACCTGTAGTGAAGGAGCTAGTGAAGTGTCTGTCTGCTG 7598
Db 9010 CTCAGGAGCGGTCTTGTCTACCTGTAGTGAAGGAGCTAGTGAAGTGTCTGTCTGCTG 9069
QY 7599 TCAATGTCTATATGTCGACAGGCGCTCTGATCAAGCCATGCGCTGCGGAGAAAGTAA 7658
Db 9070 TCGATGTCTTACATGAGGACCGCTCTGATCAGCCATGCGCTGCGGAGAAACCAAG 9129
QY 7659 CTGCCCATCAACCCCTTTCAGCAACTCTTTGCTGCTGCTACCAACAATGCTCTACGCCACA 7718
Db 9130 CTGCCCGTCAATGCACTGAGCAACTCTTTGCTGCTGCTACCAACAATGCTCTACGCCACA 9189
QY 7719 ACATCCCGGAGGCAAGCTCTCGGCAAGAAAGTGCACCTTTGACAGATTCGAACTCTG 7778
Db 9190 ACATCTCGACGGCAAGCTCTCGGCAAGAAAGTGCACCTTTGACAGATTCGAACTCTG 9249
QY 7779 GATGATCAATTAACCGGACGTTACTCAAGGAGATGAAGGAGAGGCTCCACAGTTAAGCT 7838
Db 9250 GACGACGACTACCGGAGCGTCTCAAGGAGATGAAGGAGAGGCTCCACAGTTAAGCT 9309
QY 7839 AAGCTTCTATATAGAGGAGCGCTGCAAGTGCAGCGCCCACTTCGECGCAAAATCCAAA 7898
Db 9310 AAGCTTCTATCCGTGGAGAGCGCTGTAAGTGCAGCGCCCACTTCGECGCAATCAA 9369
QY 7899 TTTGGCTAGGGGCAAGGAGCTCGGAACTCTCCAGAGGCGGCTTAACACATCCG 7958
Db 9370 TTTGGCTAGGGGCAAGGAGCTCGGAACTCTCCAGAGGCGGCTTAACACATCCG 9429
QY 7959 TCCGTGTCGGAGGACTTCTGCAAGACACTGAACACCAATGACACCACTCATGSCA 8018
Db 9430 TCCGTGTCGAAGGACTTCTGCAAGACACTGAAGACCAATGACACCACTCATGSCA 9489
QY 8019 AAAAGTGAAGTTTCTGCTCCACAGAGAGGAGCGCCGCAAGCCAGCTTCGCTTATC 8078
Db 9490 AAAAGTGAAGTTTCTGCTCCACAGAGAGGAGCGCGCCGCAAGCCAGCTTCGCTTATC 9549
QY 8079 GTATTTCCGACGCTGGAGTTCGTATGCGAGAGATGGCCCTTTAGACGCTGCTCC 8138
Db 9550 GTATTTCCGAGTTTGGGGTTCGTGTGCGAGAAATGGCCCTTTAGATGTGCTCC 9609
QY 8139 ACCCTTCTCAGGCGCTGATGGGCTCTCATACGAGTTTCAATACTCCCCCAAGCAGG 8198
Db 9610 ACCCTTCTCAGGCGCTGATGGGCTCTCATACGAGTTTCAATACTCTCTGAGCAGCGG 9669
QY 8199 GTCGAGTTCTGTGTAATACCTGGAATCAAGAAATGCGCTATGCGCTTCTCATATGAC 8258
Db 9670 GTCGAGTTCTGTGTAATGCTGGAAGCGAAGAAATGCGCTATGCGCTTCTCATATGAC 9729
QY 8259 ACCCGCTGTTTGTGACTCAACGGTCACTGAGATGACATTCGTGTGTGAGGATCAATTTAC 8318
Db 9730 ACCCGCTGTTTGTGACTCAACGGTCACTGAGATGACATTCGTGTGTGAGGATCAATTTAC 9789
QY 8319 CAATGTTGTGACTTGGCCCCCGGAGCGACAGGCCATAGGCTGCTCAGAGCGGCTT 8378
Db 9790 CAATGTTGTGACTTGGCCCCCGGAGCGACAGGCCATAGGCTGCTCAGAGCGGCTT 9849
QY 8379 TACATCGGGGCTCCCTGACTTAATCAAAAGGCGAGAACTCGCGTTATCGCGCTGCGC 8438

Db 9850 TACATCGGGGGCCCCCTGACTAATTTCTAAAGGGGAGAACTCCGGCTATCGCGGTGCGC 9903
 Qy 8439 GCAAGTGGCGGTGACAGCTAGCTGGGTAATACCCCTCAATCTTACTTGAAGCCACT 8498
 Db 9910 GCGAGCGGTGTAAGACAGCTGGGTGATACCTTCAATCTTATTTGAAGCGCCT 9969
 Qy 8499 GCAGCTGTGAGTGTCAAAAGCTCCAGGACTGCACGATGCTGTGAAGCGGAGACGCTT 8558
 Db 9970 GCGGCTGTGAGTGTGGAAGCTCCAGGACTGCACGATGCTGTGAGGAGACGCTT 10029
 Qy 8559 GTCTGTTATCTGTGAAGCGGGGAGACCCAGGAGATGCGGGGCCCTACAGCTTCAAG 8618
 Db 10030 GTCTGTTATCTGTGAAGCGGGGAGACCCAGGAGATGCGGGGCCCTTCAAG 10089
 Qy 8619 GAGCTATGACTAGTATTCGCGCCCCCCCCCGGGGATCCGCCCAACAGAAATACGACTG 8678
 Db 10090 GAGCTATGACTAGTATTCGCGCCCCCCCCCGGGGATCCGCCCAACAGAAATACGACTG 10149
 Qy 8679 GAGCTATGACTAGTATTCGCGCCCCCCCCCGGGGATCCGCCCAACAGAAATACGACTG 8738
 Db 10150 GAGTGTGATAACATCATGCTCTCCAAATGTGTGCTGCGGACGATCATCTGGCAAAAG 10209
 Qy 8739 GTATACCTACCTACCGCTGACCCACACCCCTTGCACGGGCTGGTGGGAGACGCT 8798
 Db 10210 GTGTACTATCTACCGCTGACCCACACCCCTTGGCGGGCTGGTGGGAGACGCT 10269
 Qy 8799 AGACACTCTCAATCAACTCTTGGCTAGGCAATATCATGATGCGCCCAACCTATGG 8858
 Db 10270 AGACACTCTCAATCAACTCTTGGCTAGGCAATATCATGATGCGCCCAACCTATGG 10329
 Qy 8859 GCAAGGATGATCTGATGACTCACTTTTCTCCATCTCTTCTAGCTCAAGAGCAACTGAA 8918
 Db 10330 GCAAGGATGATCTGATGACTCACTTTTCTCCATCTCTTCTAGCTCAAGAGCAACTGAA 10389
 Qy 8919 AAGCCCTGATCTGATGACTCACTTTTCTCCATCTCTTCTAGCTCAAGAGCAACTGAA 8978
 Db 10390 AAGCCCTGATGATGACTCACTTTTCTCCATCTCTTCTAGCTCAAGAGCAACTGAA 10449
 Qy 8979 CAGATCATGAAAGCACTCCATGCTCTAGGCTCTTACGCACTTACCTCCAGTCTCCAGT 9038
 Db 10450 CAGATCATGAAAGCACTCCATGCTCTAGGCTCTTACGCACTTACCTCCAGTCTCCAGT 10509
 Qy 9039 GAGATCATGAGGTGGCTTCTGCTCAGGAACTTGGGGTACACCCCTTGGCAACTGG 9098
 Db 10510 GAGATCATGAGGTGGCTTCTGCTCAGGAACTTGGGGTACACCCCTTGGCAACTGG 10569
 Qy 9099 AGACATCGGGGCGAAGTGTCCGGCTTAAGCTACTGTCCAGGGGGGAGGGCGCCACT 9158
 Db 10570 AGACATCGGGGCGAAGTGTCCGGCTTAAGCTACTGTCCAGGGGGGAGGGCGCCACT 10629
 Qy 9159 TGTGGGAGATACCTTTTAACTGGGAGTAAGGACCAAGCTTAACTCACTCCAAATCCCG 9218
 Db 10630 TGTGGGAGATACCTTTTAACTGGGAGTAAGGACCAAGCTTAACTCACTCCAAATCCCG 10689
 Qy 9219 GCGCGTCCAGCTGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9278
 Db 10690 GCTCGTCCAGGAGTGGATTTATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10749
 Qy 9279 TATCAGAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9338
 Db 10750 TATCAGAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10809
 Qy 9339 GTAGGGTAGGATTTACTGCTCCCAACCGATGAAGCGGGAGCTAACCACTCCAGGCC 9398
 Db 10810 GTAGGGTAGGATTTACTGCTCCCAACCGATGAAGCGGGAGCTAACCACTCCAGGCC 10869
 Qy 9399 TTAAGCC-----ATTTCTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 9448
 Db 10870 AATAGGCCATCTGTTTTTCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 10929
 Qy 9449 TTTTCTTTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 9508
 Db 10930 TTTTCTTTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 10989

Qy 9509 CTTAGCCCTAGTCAAGCTAGCTGTAAGGTCCTGAGCCGATGACTGACGAGTGC 9568
 Db 10990 CTTAGCCCTAGTCAAGCTAGCTGTAAGGTCCTGAGCCGATGACTGACGAGTGC 11049
 Qy 9569 TGATACTGGCCTCTCTGCAGATCATGT 9595
 Db 11050 TGATACTGGCCTCTCTGCAGATCAAGT 11076
 RESULT 11
 US-09-539-601-31
 ; Sequence 31, Application US/09539601C
 ; Patent No. 6630343
 ; GENERAL INFORMATION:
 ; APPLICANT: Bartschlag, Ralf FW
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 ; FILE REFERENCE: all sequences
 ; CURRENT APPLICATION NUMBER: US/09/539,601C
 ; CURRENT FILING DATE: 2001-08-30
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 ; EARLIER FILING DATE: 1999-04-03
 ; NUMBER OF SEQ ID NOS: 51
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 31
 ; LENGTH: 11076
 ; TYPE: DNA
 ; ORGANISM: Hepatitis C virus
 ; FEATURE:
 ; NAME/KEY: 5'UTR
 ; LOCATION: (1)..(341)
 ; OTHER INFORMATION: construct I389/Core-3'/19
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (342)..(1193)
 ; OTHER INFORMATION: hepatitis C virus core - neomycin
 ; OTHER INFORMATION: phosphotransferase fusior protein
 ; FEATURE:
 ; NAME/KEY: RBS
 ; LOCATION: (1202)..(1812)
 ; OTHER INFORMATION: internal ribosome entry site from
 ; OTHER INFORMATION: encephalomyocarditis virus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1813)..(10845)
 ; OTHER INFORMATION: hepatitis C virus polyprotein from core to
 ; OTHER INFORMATION: nonstructural protein NS5B; carries cell culture
 ; OTHER INFORMATION: adaptive mutations from clone no. 13
 ; FEATURE:
 ; NAME/KEY: 3'UTR
 ; LOCATION: (10846)..(11076)
 US-09-539-601-31

Query Match 82.3%; Score 7894.6; DB 4; Length 11076;
 Best Local Similarity 90.8%; Pred. No. 0;
 Matches 8418; Conservative 0; Mismatches 839; Indels 10; Gaps 1;
 Qy 339 ACCATGAGCAGCAATCTTAAACCTCAAGAAAAACCAACGTAACCAACCGCCGCCCA 398
 Db 1810 ACCATGGGACGATCTTAACTCTAAAGAAAAACCAACGTAACCAACCGCCGCCCA 1869
 Qy 399 CAGAGCTCAAGTTCCGGGCGGTGTGATCGTTGGTGGAGTTTACTGTTGCGCGC 458
 Db 1870 CAGAGCTCAAGTTCCGGGCGGTGTGATCGTTGGTGGAGTTTACTGTTGCGCGC 1929
 Qy 459 AGGGGCCCGAGTTGGTGTGCGCGGACTAGGAAGCTTCCGAGCGTCCGAACTCGT 518
 Db 1930 AGGGGCCCGAGTTGGTGTGCGCGGACTAGGAAGCTTCCGAGCGTCCGAACTCGT 1989
 Qy 519 GGAAGGCGCAACTATCCCAAGGCTCCGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 578
 Db 1990 GGAAGGCGCAACTATCCCAAGGCTCCGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 2049

QY 579 GGGTACCCCTGGCCCTCTATGCAATGAGGCGCTGGGTGGCGAGATGGCTCCGTGCA 638
 Db 2050 GGGTACCCCTGGCCCTCTATGCAATGAGGCGCTGGGTGGCGAGATGGCTCCGTGCA 2109
 QY 639 CCCCGGGCTCCGGCTAGTTGGGCCCCACAGAACCCCGCGTAGGTCGCTAACTTG 698
 Db 2110 CCCCGGGCTCCGGCTAGTTGGGCCCCACAGAACCCCGCGTAGGTCGCTAACTTG 2169
 QY 699 GGTAAAGTTCATGATACCCCTTACATGCGGCTTCGCGATCTCATGGGTACATTCGGCTC 758
 Db 2170 GGTAAAGTTCATGATACCCCTTACATGCGGCTTCGCGATCTCATGGGTACATTCGGCTC 2229
 QY 759 GTCCGGCCCCCTTAGGGGCGCTGCGACAGGCGCTTGGCACAGCGTTCGCGGTTCTGGAG 818
 Db 2230 GTCCGGCCCCCTTAGGGGCGCTGCGACAGGCGCTTGGCACAGCGTTCGCGGTTCTGGAG 2289
 QY 819 GACGGGTGAATATGCAACAGGAACCTTGGCGGGTTCGCTCTCTATCTTCCTCTTG 878
 Db 2290 GACGGGTGAATATGCAACAGGAACCTTGGCGGGTTCGCTCTCTATCTTCCTCTTG 2349
 QY 879 GCTCTGCTGCTCTGTTGACCATCCAGCTTCGCGTTATGAAGTGGCGAACCTGTCGGG 938
 Db 2350 GCTTTGCTGCTCTGTTGACCATCCAGCTTCGCGTTATGAAGTGGCGAACCTGTCGGG 2409
 QY 939 ATATACCATGTCCAGAACGACTGCTCCAACTCAAGCAATGTGTATGAGGACGCGACGTC 998
 Db 2410 GTGTACCATGTCCAGAACGACTGCTCCAACTCAAGCAATGTGTATGAGGACGCGACGTC 2469
 QY 999 ATCANTGCATACCCCGGGTGGCTGCTGCTGTGTCAGGAGGTAAACAGTCCCGTTCGTC 1058
 Db 2470 ATCANTGCATACCCCGGGTGGCTGCTGCTGTGTCAGGAGGTAAACAGTCCCGTTCGTC 2529
 QY 1059 GTAGCGCTCACTCCACGCTCCGCGCCAGGAATGCCAGGTCGCCCACTACGCAATACGA 1118
 Db 2530 GTAGCGCTCACTCCACGCTCCGCGCCAGGACGCTAGGTCGCCCACTACGCAATACGA 2589
 QY 1119 GCGCAGTCGACTGCTGCTGGGACGCTGCTTCTGCTCCGCTATGATGAGTGGGGAT 1178
 Db 2590 GCGCATGTCGATGCTGCTGGGCGGCTGCTCTGCTCCGCTATGATGAGTGGGAGAT 2649
 QY 1179 CTGTGGGATCTATTTCTCTCTCCAGCTGTTCACTTCTGCTCCGCTCCGCGGATGAG 1238
 Db 2650 CTGTGGGATCTATTTCTCTCTCCAGCTGTTCACTTCTGCTCCGCTCCGCGGACGAG 2709
 QY 1239 ACAGTGCAGGACTGCAACTGCTCAATCTATCCGCGCATGTATCAAGTCAACGCAATGGCT 1298
 Db 2710 ACAGTACAGGACTGCAATGCTCAATATATCCGCGCAGCTACAGGTCACCGTATGGCT 2769
 QY 1299 TGGGATATGATGAACTGGTCACTACACAGCGCTTAGTGGTGTGCGAGTTGCTCCGG 1358
 Db 2770 TGGGATATGATGAACTGGTCACTACAGCAGCGCTTAGTGGTGTGCGAGTTACTCCGG 2829
 QY 1359 ATCCACAGGCTGCTGGACATGTGGCGGGGCCACTGGGGAGTCTGGCGGGCCCTT 1418
 Db 2830 ATCCACAGGCTGCTGGATGATGGTGGCGGGGCCATTTGGGAGTCTGACGGGCGCTT 2889
 QY 1419 GCTACTATTCATGGTATGGGAACTGGGCTAAGGTTCTGATGTGGCGCTACTCTTTGCC 1478
 Db 2890 GCTACTATTCATGGTATGGGAACTGGGCTAAGGTTCTGATGTGATGCTACTCTTTGCC 2949
 QY 1479 GCGGTATGCGGGAGACCCACAGCAGGGAGGGTGGCGGCCACACCACTCCGGGTTTC 1538
 Db 2950 GCGGTATGCGGGAGACCCATGATGACAGGGGGACGATGGGCCAAAAACACCCCTCGGGAT 3009
 QY 1539 AGTCCCTTTCTCATCTCGGGGCTCTAGAAAAATCCAGCTTGTGAATACCAACGCGACG 1598
 Db 3010 AGTCCCTTTCTCATCTCGGGGCTCTAGAAAAATCCAGCTTGTGAATACCAACGCGACG 3069
 QY 1599 TGGCAGATCAACAGGACTGCCCTAAATGCAATGACTCCCTCCAACTGGGTTCTTTGCC 1658
 Db 3070 TGGCAGATCAACAGGACTGCCCTGAACTGCAATGACTCCCTCAACACTGGGTTCTTTGCC 3129
 QY 1659 GGGCTGTTTTACGCACAAAGTTCAACTGTCGGGGTCCCGAGCGCATGCCAGCTGC 1718

Db 3130 GGGCTGTTCTAGTGCACAAGTTCAACTCATCTGGATGCCAGAGCGCATGCCAGCTGC 3189
 QY 1719 GCGCCCATTTGACTGTTCCGCCAGGGGTGGGCCCATCACTATATACCTATACCTAACAGC 1778
 Db 3190 AGCCCATTCGACGCGCTTCCGCTCAGGGGTGGGGGCCATCACTTACAATAGATCACACAGC 3249
 QY 1779 TCGGATCAGAGGCGCTTATTTGCTGCGCATTAACGGCTCGACCGTGTGGTGTGCTACCCGCG 1838
 Db 3250 TCGGACCAAGAGCGCTTATTTGTTGGCATACGCAACCCCGCGCGTGGGTATGTCACCCGCG 3309
 QY 1839 TCGCAGGTGTGTGTGCTCAGTGTATTTGTTTCAACCCCAAGCCCTGTTGTGGTGGGACCAACC 1898
 Db 3310 GCGCAGGTGTGTGTGCTCAGTGTACTGCTTCAACCCCAAGCCCTGTCGTGGTGGGACGACC 3369
 QY 1899 GATCGTTCCGGTGTCCCTACGTATAGCTGGGGGGAGATGAGACAGATGCTCTCTC 1958
 Db 3370 GACCGGTTCCGGCTCCCTACGTATAGCTGGGGGGAGATGAGACAGATGCTCTCTCT 3429
 QY 1959 AACACAGCGCTCCGCCAACAGGCAACTGTTTCGCTGTACATGATGAATAGTACTGG 2018
 Db 3430 AACACAGCGCGCCGCCCAAGGCAACTGTTTGGCTGTATGATGATGAATAGTACTGG 3489
 QY 2019 TTCACTAAGACGTGCGGAGGTCCCGCTGTAACTTCGGGGGGGTGCGTAAACCGCACCTTG 2078
 Db 3490 TTCACTAAGACGTGCGGAGGTCCCGCTGTAACTTCGGGGGGGTGCGTAAACCGCACCTTG 3549
 QY 2079 ATCTGCCCCACGGACTGCTTCGGAGACGCCCGAGGCTACTTACACAAATGTGGCTGC 2138
 Db 3550 ACCTGCCCCACGGACTGCTTCGGAGACGCCCGAGGCTACTTACACAAATGTGGCTGC 3609
 QY 2139 GGGCCCTGCTTGACACCTAGGTGCTAGTACACTACCATACAGCTTTGGCACTACCCCC 2198
 Db 3610 GGGCCCTGCTTGACACCCAGATGCTTGTGTCACTACCATACAGCTTTGGCACTACCCCC 3669
 QY 2199 TGCACCTCTCAATTTTCCATCTTTAAGGTTAGGATGTATGTGGGGGGGTGAGCACAGG 2258
 Db 3670 TGCACCTCTCAATTTTCCATCTTTCAAGGTTAGGATGTATGTGGGGGGGTGAGCACAGG 3729
 QY 2259 CTCAATGCGCGATGCAATTTGACCTGCGAGGAGCGCTGTAACTTGGAGGACAGGATAGG 2318
 Db 3730 CTCGAAAGCGCATGCAATTTGACCTGCGAGGAGCGCTGTAACTTGGAGGACAGGACAGA 3789
 QY 2319 TCAGAACTCAGCCCGCTGCTGTCTACACAGAGTGGCAGATACCTGCCCTGTGCTTTC 2378
 Db 3790 TCAGAGCTTACCCCGCTGCTGTCTACACAGAGTGGCAGGATTTGCCCTGTCTTTC 3849
 QY 2379 ACCACCTTACCGGCTTTATCCACTGGTTTGTATCCATCTCCATCAGAACATCGTGACGCTG 2438
 Db 3850 ACCACCTTACCGGCTTCTGCTCACTGGTTTGTATCCATCTCCATCAGAACATCGTGACGCTG 3909
 QY 2439 CAATACCTGTAGGTTAGGTTAGGCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 2498
 Db 3910 CAATACCTGTAGGTTAGGTTAGGCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 3969
 QY 2499 CTGTGCTTTTCTCTCTCTGCGCAGACGCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2558
 Db 3970 CTGTGCTTTTCTCTCTCTGCGCAGACGCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4029
 QY 2559 CTGATAGCCCGAGGTGAGGCGCCCTTAGAGAACTTGGTGGTCTCAATTCGCGCGCTTCGCTG 2618
 Db 4030 CTGATAGCTCAAGCTGAGGCGCCCTTAGAGAACTTGGTGGTCTCAATTCGCGCGCTTCGCTG 4089
 QY 2619 GCGGAGGCGCTAGGTTATCTCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2678
 Db 4090 GCGGAGGCGCTAGGTTATCTCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4149
 QY 2679 GCGAGGCTGGCTCTGCGGCGCGCTGATGCTTTTATGCGGTATGCGCGCTCTCTCTGCTC 2738
 Db 4150 GCGAGGCTGGCTCTGCGGCGCGCTGATGCTCTTACGCGGTATGCGCGCTCTCTCTGCTC 4209
 QY 2739 CTACTGCGGTTACCAACGAGCTTACGCTTGGACCGGAGATGGGTGCTACTGCTGGG 2798

Db 4210 CTGCTGGCGTTACCAACACGAGCATACGCCATGACCGGGAGATGGCAGCATCTGTGGGA 4269
Qy 2799 GGTGGGTCTGTAGGCTCGATATCTTGAACCTGTGTCACCATCTACAAAGTCTTCTC 2858
Db 4270 GGGCGGTTTTCGTAGGCTGATCTGTGACCTTGTACCGCACTATAAGCTGTCTC 4329
Qy 2859 ACTAGGCTCATATGTTGTTTACAATCTTATCAACAGAGCGGAGGCGCACATCGAAGTG 2918
Db 4330 GCTAGGCTCATATGTTGTTTACAATCTTATCAACAGAGCGGAGGCGCACATCTGCAAGTG 4389
Qy 2919 TGGGTCCCCCGCTCAACGTTTCGGGAGCGCGATGCCATCATCTCTCACTGATGTCG 2978
Db 4390 TGGATCCCCCGCTCAACGTTTCGGGAGCGCGATGCCATCATCTCTCACTGATGTCG 4449
Qy 2979 GTTCATCCAGAGTTAAATTTTGAATCAATCAACAAATCCCTGCTGCGCATCTCGCCCGCTC 3038
Db 4450 ATCCACCCAGAGCTAATCTTTACCATCAACAAATCTTGTTCGCCATCTCGTTCATC 4509
Qy 3039 ATGCTGCTCCAGGCTGCGATCAACAGAGTGCCTTCTGTCGCGCTCAAGGCTCAT 3098
Db 4510 ATGCTGCTCCAGGCTGCGATCAACAAAGTGCCTTCTGTCGCGCTCAAGGCTCAT 4569
Qy 3099 CGTCATGCAATGTTAGTTCGGAAGTCCCGGGGTGATATGTCCTCAATGCTCTCATG 3158
Db 4570 CGTCATGCAATGTTAGTTCGGAAGTTCCTGGGGTCAATATGTCCTCAATGCTCTCATG 4629
Qy 3159 AAGCTGGCGCGCTGACAGGTACGTATCTTTATTAACCATCTTACCCCATCTCGCGGACTG 3218
Db 4630 AAGTTGGCGCGCTGACAGGTACGTATCTTTATTAACCATCTTACCCCATCTCGCGGACTG 4689
Qy 3219 GCCACCGGGCTTACGAGACCTTGGGTGCGGTGAGCCCGTCTTTCGCCCATG 3278
Db 4690 GCCACCGGGCTTACGAGACCTTGGGTGCGGTGAGCCCGTCTTTCGCCCATG 4749
Qy 3279 GAGACCAAGTTCATCACTGGGAGCGACACCGCTCGGTGCGGACATCTTCTGGT 3338
Db 4750 GAGACCAAGTTCATCACTGGGAGCGACACCGCTCGGTGCGGACATCTTCTGGT 4809
Qy 3339 CTACCCGCTCTCCGCCGGAAGGGGGAAGAGATATTTTGGACCGGCTGATAGTCTCGAA 3398
Db 4810 CTGCCCGCTCTCCGCCGGAAGGGGGAAGAGATATCTTGGACCGGCTGATAGTCTCGAA 4869
Qy 3399 GGGCAAGGTGGCACTCTTTCGCCCATCAAGCGCTCTTCCCAAAACCGCGGGGCTA 3458
Db 4870 GGGCAGGGGTGGCACTCTTTCGCCCATCAAGCGCTCTTCCCAAAACCGCGGGGCTA 4929
Qy 3459 CTTCGTTGCATCATCATAGCTCACAAGCGCGGACAAAGACAGGTGCGAAGGGGAGGT 3518
Db 4930 CTTCGTTGCATCATCATAGCTCACAAGCGCGGACAAAGACAGGTGCGAAGGGGAGGT 4989
Qy 3519 CAAGTGTGTTTACCGCAACAAATCTTCTGCGGACCTGATCAACCGGCTGTGCTGG 3578
Db 4990 CAAGTGTGTTTACCGCAACAAATCTTCTGCGGACCTGATCAACCGGCTGTGCTGG 5049
Qy 3579 ACTCTGTACATGCGGTGCTGCAAGACCTAGCCGCTCAAAAGGTGCAATCAACCCAA 3638
Db 5050 ACTCTGTATCATGTGCGGCTCAAGACCTTCCCGGCCCAAGGGGCCAATCAACCCAA 5109
Qy 3639 ATGTACACCAATGTAGACCTGGAACCTCTGCGGTGCGAGCGGCCCGCCCGGGCGCGCTCC 3698
Db 5110 ATGTACACCAATGTAGACCTGGAACCTCTGCGGTGCGAGCGGCCCGCCCGGGCGCGCTCC 5169
Qy 3699 ATGACACATGCACTGTGGCAGTTCGCACTTTTACTTGTGTCAGACATCTGTGATGTC 3758
Db 5170 TTGACACCATGCACTGTGGCAGTTCGCACTTTTACTTGTGTCAGACATCTGTGATGTC 5229
Qy 3759 ATTCCGGTGGCGCGGAGCGGACAGCGGGAGTCTACTCTCCCGCCAGCCCGCTCTCC 3818
Db 5230 ATTCCGGTGGCGCGGAGCGGACAGCGGGAGTCTACTCTCCCGCCAGCCCGCTCTCC 5289
Qy 3819 TACCTGAAAGGCTCTCTCGGTGCTTCATTTGCTTTGCCCTTCGGGGGACGCTGCGGGCTC 3878
Db 5290 TACTTGAGGGCTCTTTCGGGCGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5349

Qy 3879 TTCCGGGCTCTCTGTGTCACCCCGGGGGTTCGCAAGCGCGGTGACCTTCATACCGGTGAG 3938
Db 5350 TTTCGGGCTCTCTGTGTCACCCCGGGGGTTCGCAAGCGCGGTGACCTTTGTACCGGTGAG 5409
Qy 3939 TCTATGGAATCTACCATGCGGTCTTCGGGTCTTTCACAGACAACCTCAACCCCGCGGTGTA 3998
Db 5410 TCTATGGAATCTACCATGCGGTCTTCGGGTCTTTCACAGACAACCTCAACCCCGCGGTGTA 5469
Qy 3999 CCGCAGACATTCACAGTGGGACATCTGACGCTCTCTACTGGCAGCGGCAAGACACAAA 4058
Db 5470 CCGCAGACATTCACAGTGGGACATCTACAGCCCTCTACTGGTACGCGCAAGACACTAAG 5529
Qy 4059 GTCCCGCTCTCTATGACGCCCAAGGTACAAGTCTCTGCTCTGAACCGTCCGTTGCC 4118
Db 5530 GTCCCGCTCTCTATGACGCCCAAGGTACAAGTCTCTGCTCTGAACCGTCCGTTGCC 5589
Qy 4119 GCCACCTTAGGTTTGGGGCTATATGTCAAGGCAACGGTATCGACCTTAAACATCAGA 4178
Db 5590 GCCACCTTAGGTTTGGGGCTATATGTCTAAGGCAACGGTATCGACCTTAAACATCAGA 5649
Qy 4179 ACTGGGTAAAGACCATTACACGCGGGCTCCATTAAGTCTCTCACTTATGCAAGTTTC 4238
Db 5650 ATCGGGTAAAGACCATTACACGCGGGTCCCATCACGCTTCCACCTTATGCAAGTTTC 5709
Qy 4239 CTTCGCGACCGTGGCTTCTTGGGGCGCTATGACATCATATATGTGATGATGCCAC 4298
Db 5710 CTTCGCGACCGTGGCTTCTTGGGGCGCTATGACATCATATATGTGATGATGCCAC 5769
Qy 4299 TCAACTGACTGCACTACCATCTTGGGCACTGGCACAGTCTCTGCAACCAAGCGGAGCGGT 4358
Db 5770 TCAACTGACTGCACTACCATCTTGGGCACTGGCACAGTCTCTGCAACCAAGCGGAGCGGT 5829
Qy 4359 GGAAGCGGCTCTCTGCTGCTGCGCACCGCTACACTTCCGGGATCGGTTACCGTCCACAC 4418
Db 5830 GGAAGCGGCTCTCTGCTGCTGCGCACCGCTACACTTCCGGGATCGGTTACCGTCCACAC 5889
Qy 4419 CCCATATCAGGAATAGGCTCTTCAACAAATGGAGAGATCCCTTCTATGCAAGGCC 4478
Db 5890 CCAACATCGAGAGGTGCTCTCTGACGACTGGGAGAAATCCCTTTATGCAAGGCC 5949
Qy 4479 ATCCCATGAGGCCATCAAGGGGGGAGGACATCTCATTTTGTCCATTCGAAGAGAAA 4538
Db 5950 ATCCCATGAGGCCATCAAGGGGGGAGGACATCTCATTTTGTCCATTCGAAGAGAAA 6009
Qy 4539 TGTGACGAGCTCCGCAAGCTGACAGCCCTCGGACTGACGCTGACGCTGACATTTACCG 4598
Db 6010 TGTGACGAGCTCCGCAAGCTGACAGCCCTCGGACTGACGCTGACGCTGACATTTACCG 6069
Qy 4599 GGCCTTGTGTCCTGTCATACCGCTTATCGGAGAGCTGCTTCTGCTGGCAACAGACGCT 4658
Db 6070 GGCCTTGTGTCCTGTCATACCACTAGCGGAGAGCTCATTTCTGTCAGCAAGGACGCT 6129
Qy 4659 CTATGAGCGGTTTCAACCGGAGATTTGACTGATGATCGACTGCAATCATGTGTCAAC 4718
Db 6130 CTATGAGCGGTTTCAACCGGAGATTTGACTGATGATCGACTGCAATCATGTGTCAAC 6189
Qy 4719 CAGACAGTGCAGTTCAGCTTGGATCCCACTTCACTTGGAGACGACGACCGTGCCTCCAA 4778
Db 6190 CAGACAGTGCAGTTCAGCTTGGATCCCACTTGGAGACGACGACCGTGCCTCCAA 6249
Qy 4779 GACCGGTTTCCGCTCGCAACCGCGAGTAGAATCTGGCAGGGGTAGGAGTGCATCTAC 4838
Db 6250 GACCGGTTTCCGCTCGCAACCGCGAGTAGAATCTGGGAGGAGTGCATCTAC 6309
Qy 4839 AGTGTGTGACTCCAGAGAAAGCGGCTTGGGATGTTGATCTTCTCGGTCTCTGTGAG 4898
Db 6310 AGTGTGTGACTCCAGAGAAAGCGGCTTGGGATGTTGATCTTCTCGGTCTCTGTGAG 6369
Qy 4899 TGTATGACCGGGCTCTGCTTGTGATGAGCTCACGCCGCTGAGACCTCTCGGTAGGTTG 4958
Db 6370 TGTATGACCGGGCTCTGCTTGTGATGAGCTCACGCCGCTGAGACCTCTCGGTAGGTTG 6429

QY 4959 CCGGCTTACTAAATACACAGGGTTCGCCGTCTGCCAGGACCACTGCGAGTTCGGAG 5018
Db 6430 CCGGCTTACTAAACACACAGGGTTCGCCGTCTGCCAGGACCACTGCGAGTTCGGAG 6489
QY 5019 AGCGTCTTACAGGCTTACACACAGGATGCGGCTTCTCTGCTCCAGACTTAAACAGGCA 5078
Db 6490 AGCGTCTTACAGGCTTACACACAGGATGCGGCTTCTCTGCTCCAGACTTAAACAGGCA 6549
QY 5079 GGAGCAACTTTCCTTACCTGGTGGCATATCAAGCTTACAGTGTGGCCAGGCTCAAGCT 5138
Db 6550 GGAGCAACTTTCCTTACCTGGTGGCATATCAAGCTTACAGTGTGGCCAGGCTCAAGCT 6609
QY 5139 CCACTCCATCTGGGACCAAAATGGAAGTGTCTATACAGGCTGAAACCTTACACTGAC 5198
Db 6610 CCACTCCATCTGGGACCAAAATGGAAGTGTCTATACAGGCTGAAACCTTACACTGAC 6669
QY 5199 GGGCCAAACCCCTGCTGTATAGGCTAGGAGCGGTCCAAATGAGGTCTATCTCACAC 5258
Db 6670 GGGCCAAACCCCTGCTGTATAGGCTAGGAGCGGTCCAAATGAGGTCTATCTCACAC 6729
QY 5259 CCAATAACTTAATACATCATGGCATGCTGCGGTGAACCTGGAGTCTGCTACTAGCAC 5318
Db 6730 CCAATAACTTAATACATCATGGCATGCTGCGGTGAACCTGGAGTCTGCTACTAGCAC 6789
QY 5319 TGGGTCTGGTAGGCGGAGTCTTGGAGTCTTGGCGGCAATCTGCTGACGACAGGAGT 5378
Db 6780 TGGGTCTGGTAGGCGGAGTCTTGGAGTCTTGGCGGCAATCTGCTGACGACAGGAGT 6849
QY 5379 GTGTCATTTGGGCGAGGATCATCTGTCGGGAGGAGCTGCTGCTCAACTTCTTACATCGAG 5438
Db 6850 GTGTCATTTGGGCGAGGATCATCTGTCGGGAGGAGCTGCTGCTCAACTTCTTACATCGAG 6909
QY 5439 GTCTCTACCGAGGATTCGATGAGTGGAGAGTGTGCTGCTCAACTTCTTACATCGAG 5498
Db 6910 GTCTCTACCGAGGATTCGATGAGTGGAGAGTGTGCTGCTCAACTTCTTACATCGAG 6969
QY 5499 CAGGAAATCAGCTCGCGAGCAATCAAGCAAAAGGCGCTCGGTTGTCGAAAGGCC 5558
Db 6970 CAGGAAATCAGCTCGCGAGCAATCAAGCAAAAGGCGCTCGGTTGTCGAAAGGCC 7029
QY 5559 ACCAAGCAAGCGGAGGCTGCTGCTCGGTTGAGTCCAAAGTGGCGGACCATCGAAGCC 5618
Db 7030 ACCAAGCAAGCGGAGGCTGCTGCTCGGTTGAGTCCAAAGTGGCGGACCATCGAAGCC 7089
QY 5619 TTTCTGGGAGACACATGTGGATTTTCAACGAGGATACAGTACTAGAGGCTTATCC 5678
Db 7090 TTTCTGGGAGACATGTGGATTTTCAACGAGGATACAGTACTAGAGGCTTATCC 7149
QY 5679 ACTCTGCTGGAACCCCGGATAGCATCATGTGATGGCATTTACAGCTTCTATCACTAGC 5738
Db 7150 ACTCTGCTGGAACCCCGGATAGCATCATGTGATGGCATTTACAGCTTCTATCACTAGC 7209
QY 5739 CCGCTCACACCAAAACACCTCTGTTTACATCTTGGGGGATGAGTGGCTGCCCAA 5798
Db 7210 CCGCTCACACCAAAACACCTCTGTTTACATCTTGGGGGATGAGTGGCTGCCCAA 7269
QY 5799 CTGCTCTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5858
Db 7270 CTGCTCTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7329
QY 5859 GGCAGCATAGGCTTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5918
Db 7330 GGCAGCATAGGCTTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7389
QY 5919 GCGCGGCACTGCTGCTGCTTAAAGTCAATGAGCGGAGGCTCCCTCCACGAGGACCTG 5978
Db 7390 GCGCGGCACTGCTGCTGCTTAAAGTCAATGAGCGGAGGCTCCCTCCACGAGGACCTG 7449
QY 5979 GTCACTTACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6038
Db 7450 GTCACTTACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7509
QY 6039 GCAATACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6098

Db 7510 GCGATACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7569
QY 6099 ATAGGCTTGGCTTGGCGGGTAAACCAAGTCTCCCTAGGCTATGCTGCTGAGAGGAC 6158
Db 7570 ATAGGCTTGGCTTGGCGGGTAAACCAAGTCTCCCTAGGCTATGCTGCTGAGAGGAC 7629
QY 6159 GCTGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6218
Db 7630 GCTGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7689
QY 6219 CTCACCAAGTGAATTAATGAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6278
Db 7690 CTCACCAAGTGAATTAATGAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7749
QY 6279 GTTGGGATGGAATGATGAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6338
Db 7750 GTTGGGATGGAATGATGAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7809
QY 6339 CTGCGCGCTTACCGGGAGTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6398
Db 7810 CTGCGCGCTTACCGGGAGTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7869
QY 6399 CCGGGGAGCGGATCATGCAAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6458
Db 7870 CCGGGGAGCGGATCATGCAAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7929
QY 6459 AAAAACCGTTCATGAGGATGCTAGGGCTAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6518
Db 7930 AAAAACCGTTCATGAGGATGCTAGGGCTAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7989
QY 6519 TTCCCATCAACGATACACACGCGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6578
Db 7990 TTCCCATCAACGATACACACGCGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8049
QY 6579 AGGCGCTTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6638
Db 8050 AGGCGCTTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8109
QY 6639 CACTACGTCAGCGGATGACCACTGACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6698
Db 8110 CACTACGTCAGCGGATGACCACTGACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8169
QY 6699 GAATTTCTTACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6758
Db 8170 GAATTTCTTACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8229
QY 6759 CTCTTACGCGGAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6818
Db 8230 CTCTTACGCGGAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8289
QY 6819 CTCTTACGCGGAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6878
Db 8290 CTCTTACGCGGAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8349
QY 6879 CACATTAAGCAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6938
Db 8350 CACATTAAGCAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8409
QY 6939 AGCTCATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6998
Db 8410 AGCTCATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8469
QY 6999 GACTTCCCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7058
Db 8470 GACTTCCCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8529
QY 7059 AACTTACTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7118
Db 8530 AACTTACTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8589
QY 7119 CACGCGGAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7178

QY 9339 GTAGGGTAGGATTTACCTGCTCCCAACCGATGAACGGGAGCTAACCACTCCAGGCC 9398
Db 10810 GTAGGGTAGGATCTATCTACTCCCAACCGATGAACGGGAGCTAACCACTCCAGGCC 10869
QY 9399 TTAAGCC-----ATTTCCTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 9448
Db 10870 AATAGGCCATCGTGTGTTTTTCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 10929
QY 9449 TTTTCTTTCT 9508
Db 10930 TTTTCTTTCT 10989
QY 9509 CTTAGCCCTAGTCACCGCTAGCTGTGAAGGTCCTGTGACCGCATGACGTGACAGATGC 9568
Db 10990 CTTAGCCCTAGTCACCGCTAGCTGTGAAGGTCCTGTGACCGCATGACGTGACAGATGC 11049
QY 9569 TGATACGTGGCTCTCTGACATCATGT 9595
Db 11050 TGTACTGACCTCTCTGACATCATGT 11076

RESULT 12

US-08-324-977-13
; Sequence 13, Application US/08324977
; Patent No. 5747339
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Armstrong, Westerman, Hattori, Mclelland &
; ADDRESSES: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Stevens-Smith, Theresa M.
; REGISTRATION NUMBER: 36,281
; REFERENCE/DOCKET NUMBER: 900703D
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9030 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from genomic RNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..9030
; OTHER INFORMATION: /note: "sequence = 333 - 9362 of
; OTHER INFORMATION: SEQ ID NO: 1"
; FEATURES:
; NAME/KEY: CDS
; LOCATION: 1..9030
; US-08-324-977-13
Query Match 81.4%; Score 7809.2; DB 1; Length 9030;
Best Local Similarity 91.6%; Pred. No. 0;
Matches 8267; Conservative 0; Mismatches 763; Indels 0; Gaps 0;
QY 342 ATGAGCACGAATCTTAACCTCAAGAAAAAACAACGTAACCAACCGCGCCACAG 401
Db 1 ATGAGCACGAATCTTAACCTCAAGAAAAAACAACGTAACCAACCGCGCCACAG 60
QY 402 GACGTCAAGTTCGCGGCGGTGTCAGATCGTTGGTGGAGTTTACCTTTGCCGCGCAGG 461
Db 61 GACGTCAAGTTCGCGGCGGTGTCAGATCGTTGGTGGAGTTTACCTTTGCCGCGCAGG 120
QY 462 GCGCCCAAGTTGGGTGTCGCGCGACTAGGAGGTTCCGAGCGTCCGACACTCGTGGGA 521
Db 121 GCGCCCAAGTTGGGTGTCGCGCGCCAGGAGACTTCCGAGCGTCCGACACTCGTGGGA 180
QY 522 AGGCGACAACCTATCCAAAGGCTCGCGACCCGAGGCGCAGGGCTTGGGCTCAGCCCGGG 581
Db 181 AGGCGACAACCTATCCCAAGGCTCGCGCGCCGAGGCGCAGGACCTTGGGCTCAGCCCGGG 240
QY 582 TACCTTTGGCCCTCTATGGCAATGAGGCGCTGGGGTGGGAGAGTGGCTCTGTCAACC 641
Db 241 TACCTTTGGCCCTCTATGGCAATGAGGCGCTTAGGGTGGGAGAGTGGCTCTGTCAACC 300
QY 642 CGCGCTCCCGCCCTAGTTGGGCGCCACGACCCCGCGCTAGGTCCGCTAACCTTGGT 701
Db 301 CGCGCTCCCGCCCTAGTTGGGCGCCACGACCCCGCGCTAGGTCCGCTAACCTTGGT 360
QY 702 AAGGTATCATGATACCTTACATGCGGCTTCGCGATCTCATGGGTACATTCGCTCGTC 761
Db 361 AAGGTATCATGATACCTTACATGCGGCTTCGCGATCTCATGGGTACATTCGCTCGTC 420
QY 762 GCGCCCCCTTAGGGGGCGCTGCGAGGCGCTTGGCACACGCTGCGGTTCTGGAGAC 821
Db 421 GCGCCCCCTTAGGGGGCGCTGCGAGGCGCTTGGCACACGCTGCGGTTCTGGAGAC 480
QY 822 GCGGTGAATATGCAACAGGGAACCTTGGCCCGGTTGCTCTTCTCTATCTTCTCTTGGCT 881
Db 481 GCGGTGAATATGCAACAGGGAACCTTGGCCCGGTTGCTCTTCTCTATCTTCTCTTGGCT 540
QY 882 CTGCTGCTCTGTTGACCATCCAGCTTCGCTTATGAGTGCACAGCTGTCGCGGATA 941
Db 541 CTGCTGCTCTGTTGACCATCCAGCTTCGCTTATGAGTGCACAGCTGTCGCGGATA 600
QY 942 TACATGTGCAAGACGATGCTCCAACTCAAGCATTTGATGAGGACGCGGAGTATC 1001
Db 601 TATCATGTGCAAGACGATGCTCCAACTCAAGCATTTGATGAGGACGCGGAGTATC 660
QY 1002 ATGATACTCCCGGCTGGTCCCTGTGTTCCAGAGGCTTAACAGCTCCCGTTGTGGGTA 1061
Db 661 ATGATACTCCCGGCTGGTCCCTGTGTTCCAGAGGCTTAACAGCTCCCGTTGTGGGTA 720
QY 1062 GCGTCACTCCACGCTCGCGGCGAGGAATGCCAGCTGCCCACTAGACAATACGAGC 1121

Db	721	CGGTCCTCCACCGCTCGACCGAGAAAGTCACCATCCCAACGACATACGACGC	780
Qy	1122	CACGTGACATTGCTTCGTTGGGACCGCTGCTTTCTGCTCCGCTATGTACGTGGGGGATCTC	1181
Db	781	CACCTGATCTGCTCGTTGGGGGCGCTTCCTCTGTTCCGCTATGTACGTGGGGACCTC	840
Qy	1182	TGGCGATCTATTTTCCTCGTCTCCACGCTGTTACCTTCTCGCTCGCGGCGATGAGACA	1241
Db	841	TGCGGATCTGTTTTCCCTCGTCTCTCAGCTGTTCACCTTCTCGCTCGCGGCGATGAGACA	900
Qy	1242	GTGCAGGACTGCAACTGCTCAATCTATCCCGGCCATGTATCAGCTACCGCATGGCTTGG	1301
Db	901	TTACAGGACTGTAAGTCTCAATTTATCCCGGCCATGTGTGGGTCAACGTATGGCTTGG	960
Qy	1302	GATATGATGATGAAGTCTCACTTACAAACAGCCCTAGTGGTGTGCGAGTTGCTCCGGATC	1361
Db	961	GACATGATGATGAAGTCTGCTCCCCACAAACAGCCCTAGTGGTGTGCGAGTTGCTCCGGATC	1020
Qy	1362	CCACAAGCTGCTGTGACATGCTGCGCGGGGCCACATGGGAGTCCNMGCGGGCCTTGCC	1421
Db	1021	CCACAAGCCGCTGTGGACATGGTGGCGGGGCCACATGGGAGTCTTGGCGGGCCTTGCC	1080
Qy	1422	TACTATTCCATGGTAGGAACTGGGCTAAGTCTTCTGATTGTGGCGTACTCTTTGCCCGC	1481
Db	1081	TACTATTCCATGGCGGGAACTGGGCTAAGTCTTCTGATTGTGATGTACTTTTGTCTGCG	1140
Qy	1482	GTTTACCGGGAGACCCACACGACGGGGAGGTTGGCCGGCCACACCACTCTGGGTTTCAG	1541
Db	1141	GTTTACCGGGGATACCCACGTCGACAGGGGGGGCGAAGCCAAAACCAACACAGGCTCGTG	1200
Qy	1542	TCCCTTTTTCATCTGGGGGCGTCTCAGAAAAATCCAGCTTGTGAATACCAACGGCAGCTGG	1601
Db	1201	TCCATGTTCCAGCTGGCGCGTCTCAGAAAAATCCAGCTTATAAACAATGGAGTTGG	1260
Qy	1602	CACATCAACAGGACTGCGCTAAATTGCAATGCACTCCCTCCAACTGGGTTCTTTGCCCGC	1661
Db	1261	CACATCAACAGGACTGCGCTGAACTGCAATGCACTCTCTCCAGATGGGTTCTTTGCCCGC	1320
Qy	1662	CTGTTTTTACGCACAAAGTTCAACTCTGTCGGGTGCCGGAGCGCATGGCCAGCTGCCGC	1721
Db	1321	CTGTTCTACACANATGTTTCAACTCTGTCGGGTGCCAGGGCGCATGGCCAGTGGCCGC	1380
Qy	1722	CCCATTGACTGGTTTCGCCCGAGGGTGGGGCCCAATCACTTACTAAGCCTTAACAGCTCG	1781
Db	1381	ACCATTCACAAGTTTCGACCAAGGATGGGTGCCCAATTACTTATGTGAGTCTAGCAGATCA	1440
Qy	1782	GATCAGAGGCTTATTCTGTGCATTACGCGCTCGACCGTGTGTGTGTCGTACCGCGTGC	1841
Db	1441	GACCAGAGGCCATTGCTGGCACTACCACTCCCAATGTACATCGTACTGGGTGTCG	1500
Qy	1842	CAGTGTGTGGTTCAGTGTATTGTTTCACCCCAAGCCCTGTGTGTGGGGACACCCGAT	1901
Db	1501	GAGTGTGTGGGCCCAGTGTACTGTTTCACCCCAAGCCCTGTGTGTGGGGACGACCCGAT	1560
Qy	1902	CGTTCGGGTGTCCTACGTATAGTGGGGAGATGAGACAGCTGTATGCTCTCTCAAC	1961
Db	1561	CGTTTCGGTGTCCCTACGTATAGTGGGGAGAACAGACTGACGTGCTGCTGCTCAAC	1620
Qy	1962	AACACGGCTCCGCCACAAAGGCATGTTGTGGGTGTATCATGGATGAATAGTACTGGGTTTC	2021
Db	1621	AACACGGCGCGCGCAAGGCAACTGTTGTGGGTGTACATGGATGAATAGTACTGGGTTTC	1680
Qy	2022	ACTTAAGACGTGCGAGGTCCCCCGTTAACTNCGGGGGGTTCGGTAAACCGGACCTTGATC	2081
Db	1681	ACCAAGACATGTGGGGGGCCCCGTGTAACTTCGGGGGGTTCGGGCAACAACCTTGACC	1740
Qy	2082	TGCCCCACGGAAGCTCTTTCCGGAAGACCCCGAGGCTACTTACACAAATGTGGCTCGGG	2141
Db	1741	TGCCCCACGGAAGCTCTTTCCGGAAGACCCCGAGGCTACTTACACAAATGTGGTTTCGGG	1800
Qy	2142	CCCTGGTTTGACCTAGGTGCTTAGTAGACTCCCATCAGGCTTTTGGCACTACCCCTGC	2201

Db	1801	CCITGGCGTGACACCTAGTGCATGGTTGACATATCCATACAGGCTCTTGGCAATTACCCCTCGC	1861
Qy	2202	ACTCTCAATTTTTCACATCTTTAAAGTTAGGATGTATGTGGGGGCGCTGGAGCACAGGCTC	2261
Db	1861	ACTGTTAACTTTACCACTCTTCAAGGTTAGGATGTATGTGGGGGGGTGGAGCACAGGCTC	1920
Qy	2262	AATGCCGATGCAAATTGGACTCGAGGAGCGCTGTAACTTTGAGAGCACGGGATAGGTCA	2321
Db	1921	AATGCTGCATGCAAATTGGACCCGAGAGAGCGTTGTGACTTGGAGACACAGGATAGCCGG	1980
Qy	2322	GAACTCAGCCCGCTGCTGCTCTCAACAAGAGTGGCAGATACTGCCCTGTGCTTTCAAC	2381
Db	1981	GAGCTCAGCCCGCTGCTGCTCTCAACAGAGTGGCAGTACTGCGCCTGTTCTTTCACC	2040
Qy	2382	ACCCTACGGGCTTATCCACTGTTTGATCCATCTCATCAGAAACATCGTGGACGTGCAA	2441
Db	2041	ACCCTACAGCTCTGTCTCACTGGCTTGAATCACTCCATCAGAAACATCGTGGACGTGCAA	2100
Qy	2442	TACCTGTACAGTGTAGGCTCAGAGTGTGTCTCTCTTTTGCATCAAAATGGGAGTACATCC	2501
Db	2101	TACCTATACGTTAAGGTCAGCGGTTGTCTCTTTTGCAATCAAAATGGGAGTATGCTCTG	2160
Qy	2502	TTGCTTTTCTCTCTCTGGCAGACCGCGGCTGTGTGCTCTTGTGCATCAAAATGGGAGTACATCC	2561
Db	2161	TTGCTTTTCTCTCTCTAGCGACGCACTGTGTGTGCTCTGTGTGATGATGCTGCTG	2220
Qy	2562	ATAGCCAGAGCTGAGCGCGCTTAGAGAACTTGGTGGTCTCTCAATGGCGCGTCCGTGGCC	2621
Db	2221	ATAGCCAGCGCCGAGCGCGCTTGGAGAACTGCTGGTCTCTCAATGGCGCGTCTGTGGCC	2280
Qy	2622	GGAGCGCATGCTATCTCTCTTCTTCTTGTGTTCTTCTGCGCGCCCTGGTACATTAAGGGC	2681
Db	2281	GGCGCATGGCATCCTCTCTCTTCTTGTGTTCTTCTGTGCGCGCTGTACATCAAGAGC	2340
Qy	2682	AGGCTGCTCTCTGGGCGGCGGTATGCTTTTATTTAGGCGTATGGCGGTCTCTCTGCTCCTA	2741
Db	2341	AGGCTGCTCTCTGGGCGGCGCATATGCTCTTTATTTGGGCTGTGGCGGTCTCTCTGCTCTG	2400
Qy	2742	CTGGCGTTACACACAGAGCTTAGCGCTTGGACCGGGAGATGGCTGATCGTGGCGGGGT	2801
Db	2401	CTGGCATACACCGGAGCTTAGCGCATAGACCGGGAGATGGCTGATCGTGGGAGGC	2460
Qy	2802	CGGTTCTTGTAGTCTGTGATTTCTTGACCTTGTCAACATACAAAGTGTTCCTCACT	2861
Db	2461	CGGTTTCTTGTGGCTGTGGTACTCTGACTTGTCAACATACAAAGTGTTCCTCGCT	2520
Qy	2862	AGGCTCATATGTTGTTACATAACTTTATCACAGAGCGGAGCGGACATCGCAAGTGTGG	2921
Db	2521	AGGCTCATATGTTGTTACATAATTTTACACAGAGCGGAGGAGTCTACATGTGTGTGG	2580
Qy	2922	GTCCCCCCTCAACGTTGGGAGGCGCGATGCCATCATCTCTCTCATCGTGTGGGTT	2981
Db	2581	ATCCCCCCTCAACGTTGGGAGGCGCGATGCCATCATCTCTCTCATGTGGCGATC	2640
Qy	2982	CATCCAGAGTTAATTTTGTACATCAACAACTCTGCTCGCATCTCTCGCCCGCTCATG	3041
Db	2641	CATCCAGAGTTAATTTTGTACATCAACAACTCTTAAATGCCATATCTGCTCGCTCATG	2700
Qy	3042	GTGCTCCAGGCTGGCATAAACGAGTGGCCGTACTTCTGTGGCGCTCAAGGGCTCATTCGT	3101
Db	2701	GTGCTCCAGGCTGGCATAAACGAGTGGCCGTACTTCTGTGGCGCTCAAGGGCTCATTCAT	2760
Qy	3102	GCATGCATTTAGTGGAAAAGTTCGCGGGGTCATTATGTCAAATGGTCTTCATGAAG	3161
Db	2761	GCATGCATTTAGTGGAAAAGTTCGCTGGGGGTCAATTATGTCCAAATGGGCTTCATGAAG	2820
Qy	3162	CTGGGCGCGTGCACAGTACGTACGTTTATAACAATCTTACCCCACTGCGGAGCTGGGCC	3221
Db	2821	CTGGGCGCGTGCACAGCAGCTACATTATCAACAATCTTACCCCGCTACGGGATTTGGCCA	2880
Qy	3222	CACGCGGCGCTACAGACCTTTCGGTGGCGGTAGAGCCGCTGTCTCTCCGCATGGAG	3281
Db	2881	CGCGCGGCGCTACAGACCTTTCGGTGGCGGTAGTGGAGCGCTGTCTCTCCGACATGGAG	2940

QY 3282 ACCAAGTTCATCACCTGGGAGCAGACACCGGTGCTGTGGGACATCATCTTGGTCTA 3341
 DB 2941 ACCAAGATCATCACTGGGAGCAGACACCGGGGTGTGGGACATCATCTTGGTCTG 3000
 QY 3342 CCCGTCTCCGCGGAGGAGGAGAGATATTTTGGGACCGGTGTAGTCTCGAAGG 3401
 DB 3001 CCCGTCTCCGCGGAGGAGGAGAGATCTCCCGGGCCGGCCGAGTCTTGAAGG 3060
 QY 3402 CAAGGGTGGGACTCCTTGGGCCATCAGCGCTACTCCCAACAAACCGGGGCTACTT 3461
 DB 3061 CGGGGTTGGGACTCCTCGGCCCATCAGGCTTACTCCCAACAGACCGGGGCTACTT 3120
 QY 3462 GGTTCATCATCACTAGCCTCACAGGCGGGACAGAAACCAAGGTCCAAAGGGAGGTTCAA 3521
 DB 3121 GGTTCATCATCACTAGCCTTACAGGCGGGACAGAACCAAGTTCAGAGGAGGTTTCA 3180
 QY 3522 GTGGTTTACGGACACATCTTCTTGGGACCTGCATCAACGGGCTGTGTGGACT 3581
 DB 3181 GTGGTTTCCACCGCAACAATCTTCTTGGGACCTGCATCAACGGGCTGTGTGGACC 3240
 QY 3582 GTCTACCATGGGCTGGCTCGAAGACCTAGCGGTCCAAAGGTCCTCAATCACCACAAATG 3641
 DB 3241 GTTACCATGTGTGTGTCTCAGACCTTACCGCGCCAAAGGGCCCAATCACCAGATG 3300
 QY 3642 TACACCAATGTAGACCTCGTCTGGCTGGCAGGCGCCCGCGGGCGGCTCCATG 3701
 DB 3301 TACACTAATGTGGACGACCTCGTCTGGCTGGCCCAAGCCCCCGGGCGGCTTCCCTTG 3360
 QY 3702 ACACCATGACCTGTGGGAGCTCGGACCTTACTTGTGTGTGTGTGTGTGTGTGTGTGT 3761
 DB 3361 ACACCATGACCTGTGGGAGCTCGGACCTTACTTGTGTGTGTGTGTGTGTGTGTGTGT 3420
 QY 3762 CCGGTGCGCGGAGGCGACAGACAGGGAAGTCTACTCTCCCGAGGCGGCTGTCTTAC 3821
 DB 3421 CCGGTGCGCGGAGGCGACAGTAGGGGAGGCTCTCTCCCGAGGCGGCTGTCTTAC 3480
 QY 3822 CTGAAAGGCTCCTCGGT 3881
 DB 3481 TGAAGGCTCTTGGGT 3540
 QY 3882 CGGCTGT 3941
 DB 3541 CGGCTGT 3600
 QY 3942 ATGGAATACATCGGCTCTCGGTCTTTCAGACAACTCAACCCCGCGGCTGTACCG 4001
 DB 3601 ATGGAATACTATCGGCTCTCGGTCTTTCAGGACACTCATCTCCCGCGGCTGTACCG 3660
 QY 4002 CAGACATTCGAAGTGACATCTGTGACGCTCTTACGACGCGCAAGACACCAAGTG 4061
 DB 3661 CAGTCAATTCGAAGTGACATCTGTGACGCTCTTACGACGCGCAAGAGTACTAAGTG 3720
 QY 4062 CCGCTGT 4121
 DB 3721 CCGCTGT 3780
 QY 4122 ACCTTAGGGTTTGGGGGTATGTGTCAAGGACACAGGTATCGCACCTTAAATCAGAACT 4181
 DB 3781 ACCTTAGGGTTTGGGGGTATGTGTCAAGGACACAGGTATTTGACCCCAATCAGAACT 3840
 QY 4182 GGGGTAGGACCATTTACCGGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4241
 DB 3841 GGGGTAGGACCATTTACCGAGGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3900
 QY 4242 GCCGAGGT 4301
 DB 3901 GCCGATGT 3960
 QY 4302 ACTGACTGACTACCATCTTGGGAGCTCGGACAGTCTGTGGACCAAGCGGAGCGGCTTGA 4361
 DB 3961 ACTGACTGACTACCATCTTGGGAGCTCGGACAGTCTGTGGACCAAGCGGAGCGGCTTGA 4020

QY 4362 GCGGGCTCTGCTGTCTGCGCACCGCTACACTCTCGGATCGGTACCGTCCACACCCC 4421
 DB 4021 GCGGGCTTGTCTGTCTGCGCACCGCTACCGCTCCGGATCGGTACCGTCCACACCCA 4080
 QY 4422 AATATCGAGGAAATAGGCTGTCCAAATGGAGAGATCCCTTCTTATGGCAAGCCATC 4481
 DB 4081 AACATCGAGGAGGTGGCTGTCTAATATCTGGAGAGATCCCTTCTTATGGCAAGCCATC 4140
 QY 4482 CCCATTGAGGCAATCAAGGGGGGAGGAGCTCTCTATTTCTGCCATTCAGAGAAATGT 4541
 DB 4141 CCCATTGAGGCAATCAGGGGGGAGGAGCTCTCTATTTCTGTCAATCCAGAGAGTGC 4200
 QY 4542 GACGAGCTCGCGCAAAAGCTGACAGGCTCGGACTGAACGCTGTAGCATATTTACCGGGC 4601
 DB 4201 GACGAGCTCGCGCAAAAGCTGTCAAGGCTCGGATCAAGCTGTGGCGTATTTACCGGGG 4260
 QY 4602 CTTGATGTTCGCTCATACCGCTTATCGGAGACCTGTGTGTGTGTGTGTGTGTGTGTGT 4661
 DB 4261 CTTGATGTTCGCTCATACCAACTATCGGAGACCTGTGTGTGTGTGTGTGTGTGTGTGT 4320
 QY 4662 ATGACGGGTTTACCGGCGATTTTGTACTCAGTGTATCGACTGCAATACATGTGTCAACCCAG 4721
 DB 4321 ATGACGGGTTTACCGGCGATTTTGTACTCAGTGTATCGACTGCAATACATGTGTCAACCCAG 4380
 QY 4722 ACAGTGCATTTACGTTTGTATCCACCTTCAACATTCAGAGACGACGCTGCCCAAGAC 4781
 DB 4381 ACAGTGCATTTACGTTTGTATCCACCTTCAACATTCAGAGACGACGCTGCCCAAGAC 4440
 QY 4782 GCGGTGCGGCTCGCAACCGGCGAGGTAGAACTGCGAGGGGTAGGAGTGGCATCTACAGG 4841
 DB 4441 GCACTGCGGCTCGCAACCGGCGAGGTAGAACTGCGAGGGGTAGGAGTGGCATCTACAGG 4500
 QY 4842 TTTGTGATTCAGAGAGACGCGCTCGGCGATGTTCGATTCCTTCGCTGTGTGTGTGTGTGT 4901
 DB 4501 TTTGTGATTCAGAGAGACGCGCTCGGCGATGTTCGATTCCTTCGCTGTGTGTGTGTGTGT 4560
 QY 4902 TATGACCGGCTGT 4961
 DB 4561 TATGACCGGCTGT 4620
 QY 4962 GCTTACCTAATAACACAGAGGTTGCCGTGTGCGAGGACCATCTGAGAGTCTTGGGAGAGC 5021
 DB 4621 GCCTTACCTAATAACACAGAGGTTGCCGTGTGCGAGGACCATCTGAGAGTCTTGGGAGAGT 4680
 QY 5022 GTCTTACAGGCTCACCCACATAGATGCGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5081
 DB 4681 GTCTTACAGGCTCACCCACATAGATGCGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4740
 QY 5082 GACAACTTTCTTACTTGT 5141
 DB 4741 GACAACTTTCTTACTTGT 4800
 QY 5142 CTTCCATCTGTGGGACCAATGTGGAAGTGTCTCATAGGCTGTAAACCTACATCTGACGGG 5201
 DB 4801 CTTCCATCTGTGGGACCAATGTGGAAGTGTCTCATAGGCTGTAAACCTACATCTGACGGG 4860
 QY 5202 CCAACACCCCTGT 5261
 DB 4861 CCAACACCCCTGT 4920
 QY 5262 ATAACTAAATACATCAATGTGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5321
 DB 4921 ATAACTAAATACATCAATGT 4980
 QY 5322 GTCTGT 5381
 DB 4981 GTCTGT 5040
 QY 5382 GTCAATGT 5441
 DB 5041 GTCAATGT 5100
 QY 5442 CTCTACAGGAGTTCGATGAGATGGAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5501

Db 5101 CTTACAGGAGTTTCGATGAATGAGAGTGCCTCGCACCTCCCTTACATGAGCAG 5150
QY 5502 GGAATGAGCTCGCGAGCAATTAAGCAAAAGGCGTCGGGTGTGTGCAAAAGCCGAC 5561
Db 5161 GGAATGAGCTCGCGAGCAATTAAGCAAAAGGCGTCGGGTGTGTGCAAAAGCCGAC 5220
QY 5562 AAGCAAGCGGAGGTGTGCTCCGCTGTGAGTCCAAGTGGCGAGCCCTTGAGACCTTC 5621
Db 5221 AAAAAGCGAGGTGTGCTCCGCTGTGAGTCCAAGTGGCGAGCCCTTGAGACATTC 5280
QY 5622 TGGCGAAGCAATGCAATTTTCATCAGCGAATACAGTACCTAGCAGGCTTATCCACT 5681
Db 5281 TGGCGAAGCAATGCAATTTTCATCAGCGGATACAGTACCTAGCAGGCTTATCCACT 5340
QY 5682 CTGCTTGAAACCCCGGAGTACATGATGCAATTTACAGCTTCTATCAGTAGCCCG 5741
Db 5341 CTGCTTGAAACCCCGGAGTACATGATGCAATTTACAGCTTCTATCAGTAGCCCG 5400
QY 5742 CTCACAGCCHAAACACCTCCTCTGTAAACATCTGGGGGAGTGGGTGGCCAACTC 5801
Db 5401 CTCACAGCCHAAAGTACCTCCTCTGTAAACATCTGGGGGAGTGGGTGGCCAACTC 5460
QY 5802 GCTCCTCCAGCGCTGCTCAGCTTTCGTTGGGCGCGCATCGCGGAGCGGCTGTGSC 5861
Db 5461 GCTCCTCCAGCGCTGCTCAGCTTTCGTTGGGCGCGCATCGCGGAGCGGCTGTGSC 5520
QY 5862 AGCATAGCCTTGGAAAGTCTCTGAGACATCTTGGCGGCTATGGGCGAGGGGTAGCC 5921
Db 5521 AGCATAGCCTTGGAAAGTCTCTGAGACATCTTGGCGGCTATGGGCGAGGGGTAGCC 5980
QY 5922 GGCACATCTGGGCTTAAAGTATAGAGCGGAGGTGCCCTCCACGAGGACCTGCTC 5981
Db 5581 GGCACATCTGGGCTTAAAGTATAGAGCGGAGGTGCCCTCCACGAGGACCTGCTC 5640
QY 5982 AACTTACTCCTGCTCCT 6041
Db 5641 AACTTACTCCTGCTCCT 5700
QY 6042 ATACTGCTCGGACGTTGGGCGCGGAGAGGGGCTGTGAGTGGATGAACCGGCTGATA 6101
Db 5701 ATACTGCTCGGACGTTGGGCGCGGAGAGGGGCTGTGAGTGGATGAACCGGCTGATA 5760
QY 6102 GCGTTCGCTTCGGGGTAACAGCTCTCCCTACGCACTATGCTCTGAGAGGACGCT 6161
Db 5761 GCGTTCGCTTCGGGGTAACAGCTCTCCCTACGCACTATGCTCTGAGAGGACGCT 5820
QY 6162 GCAGCAGTGTCACTCAGATCCTCTCTAGCTTACCATCACTCAACCTGCTGAGCGGCTC 6221
Db 5821 GCAGCAGTGTCACTCAGATCCTCTCTAGCTTACCATCACTCAACCTGCTGAGCGGCTC 5880
QY 6222 CACAGTGGATTAATGAGGACGCTCTACGCCATGCTTCGGCTCGTGGCTAAGGAGTGT 6281
Db 5881 CACAGTGGATTAATGAGGACGCTCTACGCCATGCTTCGGCTCGTGGCTAAGGAGTGT 5940
QY 6282 TGGGATGAGATGCAAGGCTGAGCACTCTCAAGACCTGGCTCAGTCCAACTCCTG 6341
Db 5941 TGGGATGAGATGCAAGGCTGAGCACTCTCAAGACCTGGCTCAGTCCAACTCCTG 6000
QY 6342 CCGCGTTACCGGAGTCCCTTTCTCTCATGCCAAGCGGTAACAAGGAGTCTGSCGG 6401
Db 6001 CCGCAGTACCTGGAGTCCCTTTTCTCTGTCGCAACCGGTTACAAGGAGTCTGSCGG 6060
QY 6402 GGGACCGCATCATGAAAAACCTGCCCCATGCGGAGCAAGATCGCCGACATGTCAAA 6461
Db 6061 GGGACCGCATCATGAAAAACCTGCCCCATGCGGAGCAAGATCGCCGACATGTCAAA 6120
QY 6462 AACGTTTCCATGAGGATCGTAGGCTTAGAACCTGAGAGCAAGTGGCGAGGACCTTC 6521
Db 6121 AACGTTTCCATGAGGATCGTAGGCTTAGAACCTGAGAGCAAGTGGCGAGGACCTTC 6180
QY 6522 CCCATCAACGATACACACGAGGACCTTGACACCTCCCGGCGCCACTATTCAGG 6581

Db 6181 CCCATCAACGATACACACGAGGCGCTGACACCTCTCTCCAGCGCAAACTATTCAGG 6240
QY 6582 GCGCTATGGCGGCTGCTGAGGAGTACGTGGAGTTACGGGTGAGGGGATTTCCAC 6641
Db 6241 GCGCTATGGCGGCTGCTGAGGAGTACGTGGAGTTACGGGTGAGGGGATTTCCAC 6300
QY 6642 TAGCTACGCGGATGACCACTGACAAAGTAAAGTGCCTATGCCAGGTTCCGGGCCCCGAA 6701
Db 6301 TAGCTACGCGGATGACCACTGACAAAGTAAAGTGCCTATGCCAGGTTCCGGCTCTGAA 6360
QY 6702 TTCTTTCAGGAGTGAATGAGTGGGTGCAAGGTACGCTCCGGGTGCAAACTCTTT 6761
Db 6361 TTCTTTCAGGAGTGAATGAGTGGGTGCAAGGTACGCTCCGGGTGCAAGGCTCTCT 6420
QY 6762 CTACGAGGACCTCAGTTCCAGTTCGGGCTCAACCAATACCTTGTTCGGTTCGAGCTC 6821
Db 6421 CTACGAGGAGGTTTACATTCAGGTCGGGCTCAACCAATACCTTGTTCGGTTCGAGCTC 6480
QY 6822 CCATGCGAGCCCGAAACCGGACGTAAAGTGTCTTCCATGCTCACCGATCCCTCCAC 6881
Db 6481 CCATGCGAGCCCGAAACCGGATGTAGCAGTGTCTTCCATGCTCACCGATCCCTCCAC 6540
QY 6882 ATTACAGAGAGAGGCTAAGCTAGCTGGCTAGAGGCTCTCCCTCTTTTTCAGGAGC 6941
Db 6541 ATTACAGAGAGGCTAAGCTAGCTGGCTAGAGGCTCTCCCTCTTTTTCAGGAGC 6600
QY 6942 TCATCAGTACGCTGCTGCGCTCTTTTGAAGCGGACATGCACTACCCACCATGAC 7001
Db 6601 TCATCAGTACGCTGCTGCGCTCTTTTGAAGCGGACATGCACTACCCACCATGAC 6660
QY 7002 TCCCGGAGCTGACCTCATCGAGGCAACCTCTTTTGGCGGAGGAGTGGCGGAAAC 7061
Db 6661 TCCCGGAGCTGACCTCATCGAGGCAACCTCTTTTGGCGGAGGAGTGGCGGAAAC 6720
QY 7062 ATCACTGCGGTGAGTACAGGAATAGGTAAGTATCTTGGATCTCTTTCGAAACCGCTTCAC 7121
Db 6721 ATCACTGCGGTGAGTACAGGAATAGGTAAGTATCTTGGATCTCTTTCGAAACCGCTTCAC 6780
QY 7122 GCGGAGGAGGATGAGAGGAGATATCCGTGCGGCGGAGATCTTCGGAATTCAGGAG 7181
Db 6781 GCGGAGGAGGATGAGAGGAGATATCCGTGCGGCGGAGATCTTCGGAATTCAGGAG 6840
QY 7182 TTCCCTCTCAGGCTGCGGATGAGGACGCGGAGTACAATCTCTCACTGTGTAGAGTCC 7241
Db 6841 TTCCCTCTCAGGCTGCGGATGAGGACGCGGAGTACAATCTCTCACTGTGTAGAGTCC 6900
QY 7242 TGGAAAGGACCGGATGAGTCCCTCGGTGCTACAGGATGCCATTCGACATCAAG 7301
Db 6901 TGGAAAGGACCGGATGAGTCCCTCGGTGCTACAGGATGCCATTCGACATCAAG 6960
QY 7302 GCTCCTCCAATACCACTCCACGAGAAAGAGACGCTTGTCTGTACAGATCCAATGTG 7361
Db 6961 GCTCCTCCAATACCACTCCACGAGAAAGAGACGCTTGTCTGTACAGATCCAATGTG 7020
QY 7362 TCTTCTGCTTGGCGAGTTCGCACTAAGACTTCGCTGAGTTCGAGTTCGCTGCGGCTGT 7421
Db 7021 TCTTCTGCTTGGCGAGTTCGCACTAAGACTTCGCTGAGTTCGAGTTCGCTGCGGCTGT 7080
QY 7422 GATAGCGGACGCGGACGCTTCCTGAGCTTCGCTGAGTTCGAGTTCGCTGCGGCTGT 7481
Db 7081 GATAGCGGACGCGGACGCTTCCTGAGCTTCGCTGAGTTCGAGTTCGCTGCGGCTGT 7140
QY 7482 GAGTTCGCTTACT 7541
Db 7141 GAGTTCGCTTACT 7200
QY 7542 AGCGAGGCTTCTGCTTACT 7601
Db 7201 AGTACGCGTCTTGTCT 7260
QY 7602 ATGCTCTATACGTGAGAGCGGCTTCATCAAGCATGCTGCTGCGGAGGAGATGAGCTG 7661
Db 7261 ATGCTCTATACGTGAGAGCGGCTTCATCAAGCATGCTGCTGCGGAGGAGATGAGCTG 7320

7662 QY CCAATCAACCCGTTGAGCAACTCTTTGCTCCGTCACCAAAATGCTTACGCCAACA 7721
 7321 Db CCAATCAACCCGTTGAGCAACTCTTTGCTCCGTCACCAAAATGCTTACGCCAACA 7380
 7722 QY TCCGCGAGCGCAAGCTCCGGCAGAGAGAGTCACTTTGACGATGCAAGTCTGAT 7781
 7381 Db TCCGCGAGCGCAAGCTCCGGCAGAGAGAGTCACTTTGACGATGCAAGTCTGAT 7440
 7782 QY GATCAATACCGGACGCTTCAAGGAGATGAAGGCGAAGGCGTCCACAGTTAAGCTAAG 7841
 7441 Db GACCACTACCGGACGCTTCAAGGAGATGAAGGCGAAGGCGTCCACAGTTAAGCTAAG 7500
 7842 QY CTTCTATCATAGAGAGCGCTGCAAGCTGACGCGCCCAATTCGGCCCAATTCCAAATTT 7901
 7501 Db CTTCTATCATAGAGAGCGCTGCAAGCTGACGCGCCCAATTCGGCCCAATTCCAAATTT 7560
 7902 QY GGTATGCGGCAAGAGAGCTCCGGACCTATCCAGGAGGCGGTAAACACATCCGCTCC 7961
 7561 Db GGTATGCGGCAAGAGAGCTCCGGACCTATCCAGGAGGCGGTAAACACATCCGCTCC 7620
 7962 QY GTGTGGGAGGACTTGTGGAAGACACTGAAACACCAATTTGACACCAACCATCATGCAAAA 8021
 7621 Db GTGTGGGAGGACTTGTGGAAGACACTGAAACACCAATTTGACACCAACCATCATGCAAAA 7680
 8022 QY AGTGAGGTTTCTGCGTCCAAACAGAGAGGAGCGCGCAAGCTCGCTTATCGTA 8081
 7681 Db AGTGAGGTTTCTGCGTCCAAACAGAGAGGAGCGCGTAAGCGCGCTTATCGTA 7740
 8082 QY TTCCACAGACCTGGGAGTTCGTATGCGAGAGATGGCCCTTACAGCTGCTCCAC 8141
 7741 Db TTCCACAGATCTGGGAGTTCGTATGCGAGAGATGGCCCTTATGATGCTGCTCCAC 7800
 8142 QY CTTCTCAGGCGTGTGAGGCTTCTCATACGATTTCAATATCTCCCAAGCAGCGGTC 8201
 7801 Db CTTCTCAGGCTGTGAGGCTTCTCATACGATTTCAATATCTCCCAAGCAGCGGTC 7860
 8202 QY GAGTTCCTGCTGAATACCTGGAATCAAGAAATGCTTATGAGGCTTCTCATATGACACC 8261
 7861 Db GAGTTCCTGCTGAATACCTGGAATCAAGAAATGCTTATGAGGCTTCTCATATGACACC 7920
 8262 QY CGCTGTTTGTGACTCAACCGTCACTCAGAGTGCATTCGTTGAGAGTCAATTTACCAA 8321
 7921 Db CGCTGTTTGTGACTCAACCGTCACTCAGAGTGCATTCGTTGAGAGTCAATTTACCAA 7980
 8322 QY TGTGTGACTTGGCCCCCGAGCCACAGAGGCTAAGTTCGCTCAGAGCGGCTTTAC 8381
 7981 Db TGTGTGACTTGGCCCCCGAGCCACAGAGGCTAAGTTCGCTCAGAGCGGCTTTAT 8040
 8382 QY ATCGGGGTCCTGCTGACTTAATCAAGAGGCGAGAACTCGGTTATCGCGGTGCGCGCA 8441
 8041 Db ATCGGGGTCCTGCTGACTTAATCAAGAGGCGAGAACTCGGTTATCGCGGTGCGCGCG 8100
 8442 QY AGTGGCTGCTGAGCACTAGTGGGTAAATACCTTCACATGTTACTTGAAGGCCACTGCA 8501
 8101 Db AGCGCGTCTGAGCACTAGTGGGTAAATACCTTCACATGTTACTTGAAGGCCCTGCA 8160
 8502 QY GCCTGTGAGTGAAGTCCAGGACTGACAGTCTCGTGAACGAGAGACCTTGTG 8561
 8161 Db GCCTGTGAGTGAAGTCCAGGACTGACAGTCTCGTGAACGAGAGACCTTGTG 8220
 8562 QY GTTATCTGTAAGCGCGGAACTCCAGAGATGCGGCGGCTTACGAGCTTCAAGGAG 8621
 8221 Db GTTATCTGTAAGCGCGGAACTCCAGAGATGCGGCGGCTTACGAGCTTCAAGGAG 8280
 8622 QY GCTATGACTAGTATTTCCGCCCCCCCCCGGATCCGCCCAACAGATACGACTGAG 8681
 8281 Db GCTATGACTAGTATTTCCGCCCCCCCCCGGATCCGCCCAACAGATACGACTGAG 8340
 8682 QY CTGATAACATGTTCTCTCAATGTGTGCTGCGGACCGATGCTGCGCAAAAGGTA 8741
 8341 Db CTGATAACATGTTCTCTCAATGTGTGCTGCGGACCGATGCTGCGCAAAAGGTA 8400

8742 QY TACTACCTCACCGGTGACCCACACCCCTTGCACGCGGTGCGTGGGAGACAGTAGA 8801
 8401 Db TACTACCTCACCGGTGACCCACACCCCTTGCACGCGGTGCGTGGGAGACAGTAGA 8460
 8802 QY CACATCCAAATCAACTCTTGGCTAGGCAATATCATCTGATGCGCCCAACCTATGGGCA 8861
 8461 Db CACATCCAAATCAACTCTTGGCTAGGCAATATTAATGATGCGCCCACTTTGTGGGCA 8520
 8862 QY AGGATGATCTGATGACTCACTTTTTCCTCCATCTTCTAGCTCAAGAGAACTTGAAAA 8921
 8521 Db AGGATGATCTGATGACTCACTTTTTCCTCCATCTTCTAGCTCAAGAGAACTTGAAAA 8580
 8922 QY GCCCTGGATGTCAGATCTACGGGCTTGTACTCCATCTGAGGCACTTGAACCTCTCAG 8981
 8581 Db GCCCTGGATGTCAGATCTACGGGCTTGTACTCCATCTGAGGCACTTGAACCTCTCAG 8640
 8982 QY ATCAATGAACGACTCCATGGGCTTGTAGGCACTTACCTCCACAGTACTCTCCAGGTGAG 9041
 8641 Db ATCAATGAACGACTCCATGGGCTTGTAGGCACTTACCTCCACAGTACTCTCCAGGTGAG 8700
 9042 QY ATCAATGAGGTGGCTTCATGCTCAGGAACTTGGGTTACCACTTGCAGAACCTGGAGA 9101
 8701 Db ATCAATGAGGTGGCTTCATGCTCAGGAACTTGGGTTACCACTTGCAGAACCTGGAGA 8760
 9102 QY CATCGGCGCAGAGTGTCCGGCTAAGCTACTGTCCCAAGGGGAGGGCGCCACTTGT 9161
 8761 Db CATCGGCGCAGAGTGTCCGGCTAAGCTACTGTCCCAAGGGGAGGGCGCCACTTGT 8820
 9162 QY GCGAGATACCTTCTTAACTGGGCACTAAGCAAGCTTAAACTCACTCCAAATCCCGGCC 9221
 8821 Db GCGAGATACCTTCTTAACTGGGCACTAAGCAAGCTTAAACTCACTCCAAATCCCGGCC 8880
 9222 QY GCGTCCAGCTGGAGTGTCTGCGGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9281
 8881 Db GCGTCCAGCTGGAGTGTCTGCGGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8940
 9282 QY CACAGCCTGTCTGCTGCGGACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9341
 8941 Db CACAGCCTGTCTGCTGCGGACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9000
 9342 QY GGGGTAGCATTTACCTGCTCCCAACCGA 9371
 8901 Db GGGGTAGCATTTACCTGCTCCCAACCGA 9030

RESULT 13
 US-08-384-616-13
 ; Sequence 13, Application US/08384616
 ; Patent No. 5847101
 ; GENERAL INFORMATION:
 ; APPLICANT: OKAYAMA, Hiroto
 ; APPLICANT: FUKU, Isao
 ; APPLICANT: MORI, Chisato
 ; APPLICANT: TAKAMIZAWA, Akahisa
 ; APPLICANT: YOSHIDA, Iwao
 ; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
 ; NUMBER OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
 ; CORRESPONDENCE ADDRESS: 50
 ; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLend &
 ; ADDRESSEE: Naughton
 ; STREET: 1725 K St. N.W. Suite 1000
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION NUMBER: US/08/384,616


```

/ FILING DATE:
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/   APPLICATION NUMBER: US 07/769,996
/ FILING DATE: 02-OCT-1991
/ APPLICATION NUMBER: JP 2-167466
/ FILING DATE: 25-JUN-1990
/ PRIOR APPLICATION DATA:
/   APPLICATION NUMBER: JP 2-230921
/ FILING DATE: 31-AUG-1990
/ PRIOR APPLICATION DATA:
/   APPLICATION NUMBER: JP 2-305605
/ FILING DATE: 09-NOV-1990
/ PRIOR APPLICATION DATA:
/   APPLICATION NUMBER: US 07/635,451
/ FILING DATE: 28-DEC-1990
/ ATTORNEY/AGENT INFORMATION:
/   NAME: Stevens-Smith, Theresa M.
/ REGISTRATION NUMBER: 36,281
/ REFERENCE/DOCKET NUMBER: 900703B
/ TELECOMMUNICATION INFORMATION:
/   TELEPHONE: (202) 659-2930
/   TELEFAX: (202) 887-0357
/   TELEX: 440142
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 9030 base pairs
/   TYPE: nucleic acid
/   STRANDEDNESS: single
/   TOPOLOGY: linear
/ MOLECULE TYPE: cDNA from genomic RNA
/ FEATURE:
/   NAME/KEY: misc feature
/   LOCATION: 1..9030
/ OTHER INFORMATION: /note: "sequence = 333 - 9362 of
/   OTHER INFORMATION: SEQ ID NO: 1"
/ FEATURE:
/   NAME/KEY: CDS
/   LOCATION: 1..9030
/
/ US-08-384-616-13
/
/ Query Match      81.4%; Score 7809.2; DB 2; Length 9030;
/ Best Local Similarity 91.6%; Pred. No. 0;
/ Matches 8267; Conservative 0; Mismatches 763; Indels 0; Gaps 0;
/
/ Qy 342 ATGAGCAGAACTCTAAACCTCAAAGAAAAACCAAAAGTAAACCAACCGCCGCCACAG 401
/ Db 1 ATGAGCAGAACTCTAAACCTCAAAGAAAAACCAAAAGTAAACCAACCGCCGCCACAG 60
/
/ Qy 402 GACCTCAAGTTCGCGGGCGGTGCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGG 461
/ Db 61 GACCTCAAGTTCGCGGGCGGTGCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGG 120
/
/ Qy 462 GGCCCCAGGTGGTGTGCGCGGACCTAGGAAGGCTTCCGAGCGGTGCGCAACCTCGTGGGA 521
/ Db 121 GGCCCCAGGTGGTGTGCGCGGACCTAGGAAGGCTTCCGAGCGGTGCGCAACCTCGTGGGA 180
/
/ Qy 522 AGGCGACAACTATCCAAAGGCTCGCGACCCGAGGCGAGGCGCTGGGCTCAGCCCGGG 581
/ Db 181 AGGCGACAACTATCCCAAGGCTCGCGGCGCGAGGCGAGGACCTGGGCTCAGCCCGGG 240
/
/ Qy 582 TACCCTTGGCCCTCTATGGCAATGAGGCGCTGGGCTGGGCTGGGAGATGGCTCTGTCAACC 641
/ Db 241 TACCCTTGGCCCTCTATGGCAATGAGGCGCTTAGGGTGGGAGATGGCTCTGTGTCAACC 300
/
/ Qy 642 CGCGGCTCCCGGCTAGTTGGGGGCCCAACGGAACCCCGGCGTAGTTCGCTAAGCTAAGTGGGT 701
/ Db 301 CGCGGCTCCCGGCTAGTTGGGGGCCCAACGGAACCCCGGCGTAGTTCGCTAAGTGGGT 360
/
/ Qy 702 AAGTCAATCAATACCTTACATGCGGCTTCGCGCAATCTATGGGCTACATTCGCTCGTC 761
/ Db 361 AAGTCAATCAATACCTTACATGCGGCTTCGCGCAATCTATGGGCTACATTCGCTCGTC 420

```

```

/ Qy 762 GCGCCCCCTAGGGGGCGCTGCAGGGCCTTGCCACACGGTGTCCGGGTCTCGAGGAC 821
/ Db 421 GCGCCCCCTAGGGGGCGCTGCAGGGCCTTGCCACACGGTGTCCGGGTCTCGAGGAC 480
/
/ Qy 822 GCGGTGAACCTATGCAACAGGAACCTTGCOCGGTGTCTTCTCTATCTCTCTCTGGCT 881
/ Db 481 GCGGTGAACCTATGCAACAGGAACCTTGCOCGGTGTCTTCTCTATCTCTCTCTGGCT 540
/
/ Qy 882 CTGCTGTCTCTGTGTTGACCATCCAGCTTCGGCTTATGAAGTGGCCAACTGTCCGGGATA 941
/ Db 541 CTGCTGTCTCTGTGTTGACCATCCAGCTTCGGCTTATGAAGTGGCCAACTGTCCGGGATA 600
/
/ Qy 942 TACCATGTACGAAACGACTGTCTCAAACCTAAGCAATGTGTATGAGGAGCGGAGCTGATC 1001
/ Db 601 TATCATGTACGAAACGACTGTCTCAAACCTAAGCAATGTGTATGAGGAGCGGAGCTGATC 660
/
/ Qy 1002 ATGCATACTCCCGGTGCGGCGCTGCTGTTTACAGAGGTAACAGCTCCGTTCTCTGGTA 1061
/ Db 661 ATGCATACTCCCGGTGCGGCGCTGCTGTTTACAGAGGTAACAGCTCCGTTCTCTGGTA 720
/
/ Qy 1062 GCGCTCACTCCAGCGTTCGCGGCGAGGAATGCGAGGCTCCCACTACGCAATACGACGC 1121
/ Db 721 GCGCTCACTCCAGCGTTCGCGGCGAGGAATGCGAGGCTCCCACTACGCAATACGACGC 780
/
/ Qy 1122 CAGTGCAGTGTCTGTTGGGACGCGTCTTCTGCTCCGCTATGACGTTGGGGATCTC 1181
/ Db 781 CAGTGCAGTGTCTGTTGGGACGCGTCTTCTGCTCCGCTATGACGTTGGGGATCTC 840
/
/ Qy 1182 TCGGATCTATTTTCTCTCCAGCTTCGCGGCTTCGCTCTCAGCTGTTCACCTTCTCGGCA 1241
/ Db 841 TCGGATCTATTTTCTCTCCAGCTTCGCGGCTTCGCTCTCAGCTGTTCACCTTCTCGGCA 900
/
/ Qy 1242 GTCCAGGACTGCAACTGCTCAATCTATCCCGGCATCTATCAGTCAACGCAATGCTTGG 1301
/ Db 901 TTACAGGACTGTAACCTGCTCAATCTATCCCGGCATCTATCAGTCAACGCAATGCTTGG 960
/
/ Qy 1302 GATATGATGATGAATGCTGTCACCTACAAGCCCTAGTGTGTCGAGTGTCTCCGGATC 1361
/ Db 961 GATATGATGATGAATGCTGTCACCTACAAGCCCTAGTGTGTCGAGTGTCTCCGGATC 1020
/
/ Qy 1362 CCACAGCTGTCTGAGACATGTTGGGGGGGCCACTGTTGGGAGTCTTGGCGGCTTGGC 1421
/ Db 1021 CCACAGCTGTCTGAGACATGTTGGGGGGGCCACTGTTGGGAGTCTTGGCGGCTTGGC 1080
/
/ Qy 1422 TACTATTCATGTTAGGGAACCTGGGCTTAAGTTTCTGATTTGATGCTACTTTTGTGGC 1481
/ Db 1081 TACTATTCATGTTAGGGAACCTGGGCTTAAGTTTCTGATTTGATGCTACTTTTGTGGC 1140
/
/ Qy 1482 GTTCAGGGGAGACCCACAGCGGGGAGGTTGGCGGCCACACCACTCCGGTTCAAG 1541
/ Db 1141 GTTCAGGGGAGATACCCACAGCGGGGAGGTTGGCGGCCACCACTCCGGTTCAAG 1200
/
/ Qy 1542 TCCTTTTCTCTCATCTGGGGGCTCTCAGAAAAATCCAGCTTGTGAATACCAACGCGAGCTG 1601
/ Db 1201 TCCTTTTCTCTCATCTGGGGGCTCTCAGAAAAATCCAGCTTGTGAATACCAACGCGAGCTG 1260
/
/ Qy 1602 CACATCAACAGGACTGCGCTTAATGCAATGACTCCCTCCAACTGGTCTTGTTCGGCG 1661
/ Db 1261 CACATCAACAGGACTGCGCTTAATGCAATGACTCTCTCCAGACTGGTCTTGTTCGGCG 1320
/
/ Qy 1662 CTGTTTACGACACAAAGTTCAACTGTCGGGTGCGCGGAGCGGATGGCCAGCTCCGC 1721
/ Db 1321 CTGTTTACGACACAAAGTTTCAACTGTCGGGTGCGCGGAGCGGATGGCCAGCTCCGC 1380
/
/ Qy 1722 CCATTTGACTGTTTCCCGAGGGTGGGCCCCCATCACTATATCAAGCTTAACACTCG 1781
/ Db 1381 ACCATTTGACAAATTCCACAGGGATGGGTCCCATTTACTTATGCTGAGTCTAGCAGATCA 1440
/
/ Qy 1782 GATCAGAGGCTTATGCTGGCAATACGCGCTTCGAGCGCTGAGTGTGCTGCTGCTGCTG 1841
/ Db 1441 GATCAGAGGCTTATGCTGGCAATACGCGCTTCGAGCGCTTCGAGCGCTGAGTGTGCTGCTGCTG 1500
/
/ Qy 1842 CAGGTGTGTGGTCCAGTGTATTTTCACCCCAAGCCCTGTTGTGTGGGAGCACCGAT 1901

```

1501 GAGGTGTGGGCCAGTGTACTGCTTCAACCAAGCCCTGTGCTGGGACGACCGAT 1560
1502 GGTTCGGGTGCTCCCTACGATAGCTGGGGGAGATGAGACAGCTGATGCTCTCAAC 1561
1561 CGTTTCGGGTGCTCCCTACGATAGCTGGGGGAGACAGACTGATGCTGCTCAAC 1620
1562 AACAGCGGTCCGCCACCAAGGCAACTGGTTCGGCTGTACATGATGAATAGTACTGGTTC 2021
1621 AACACGCGCCCGCGCAAGGCAACTGGTTCGGCTGCACATGATGAATAGCACCGGGTTC 1680
2022 ACTAAGAGCTGGGAGTCCCGCTGTAACTACATCGGGGGGTGCTTAACCGACCTTGATC 2081
1681 ACCAAGACATGTGGGGGCGCCCGCTGTAACTACATCGGGGGGTGCGCAACACCCCTGACC 1740
2082 TGCCCCACGAGTGTCTCCGGAAGCACCCGAGGCTACTTACACAAATGTGGCTCGGGG 2141
1741 TGCCCCACGAGTGTCTCCGGAAGCACCCGAGGCTACTTACACAAATGTGGCTCGGGG 1800
2142 CCTGGTTGACACCTAGTGGCTGTAGTACTACCATACAGCTTTGGCACTACCCCTGC 2201
1801 CCTGGCTGACACCTAGTGGCTGTAGTACTATCCATACAGGCTCTGGCACTTACCCCTGC 1860
2202 ACTCTCAATTTTTCATCTTTAGGTTAGGATGATGTGGGGGGGTGGAGACAGGCTC 2261
1861 ACTGTTAATTTTACCATCTTCAAGGTTAGGATGATGTGGGGGGGTGGAGACAGGCTC 1920
2262 AATGCCGATGCAATTTGACTCGAGAGAGCGCTGTAACTTGGAGACAGGATAGGTCA 2321
1921 AATGTGTCATGCAATTTGACCCGAGAGAGCGTGTGACTTGGAGACAGGATAGGCGG 1980
2322 GAACTCAGCCCGCTGTGCTGTCTCAACAGAGTGGCAGATACTGCCCTGTCTTTACC 2381
1981 GAGCTCAGCCCGCTGTGCTGTCTCAACAGAGTGGCAGTACTGCCCTGTCTTTACC 2040
2382 ACCCTACCGGCTTTATCCACTGGTTGATCTCAATCTCCATCAGAAATCGTGAAGTGCNA 2441
2041 ACCCTACCGGCTGTGCTGTCTCAACAGAGTGGCAGTACTGCCCTGTCTTTACC 2100
2442 TACCTGTACGCTGTAGGCTCAGGTTGCTCTCTTGTCAATCAATGGGAGTACATPCCTG 2501
2101 TACCTATACGCTGTAGGCTCAGGTTGCTCTCTTGTCAATCAATGGGAGTATGCTCTG 2160
2502 TTGCTTTTCTCTCTGTCAGACGCGCGCTGTGTGCTGTGCTGTGCTGTGCTGTGCTG 2561
2161 TTGCTTTTCTCTCTGTCAGACGCGCTGTGTGCTGTGCTGTGCTGTGCTGTGCTG 2220
2562 ATAGCCAGGCTGAGCGCCCTTAGAGACTTGTGTGCTCTCAATCGCGCGCTGCTGCGC 2621
2221 ATAGCCAGGCTGAGCGCCCTTAGAGACTTGTGTGCTCTCAATCGCGCGCTGCTGCGC 2280
2622 GGAGCCCATGCTATCTCTCTCTCTTGTGTGCTCTTGTGCTGCTGCTGCTGCTGCTG 2681
2281 GCGCACATGCGATCTCTCTCTCTCTTGTGTGCTCTTGTGCTGCTGCTGCTGCTGCTG 2340
2682 AGGCTGGCTCTGCGGCGCGATGCTTTTATGGGATAGCGCGCTGCTGCTGCTGCTGCTA 2741
2341 AGGCTGGCTCTGCGGCGCGATGCTTTTATGGGATAGCGCGCTGCTGCTGCTGCTGCTG 2400
2742 CTGGGCTTACCACGAGCTTACGCTTGGACCGGGAGATGGCTGATCGTGGCGGGGT 2801
2401 CTGGGATTACCACGAGCTTACGCTTGGACCGGGAGATGGCTGATCGTGGCGGGGT 2460
2802 GCGGTTCTTGTAGGCTGTGATTTCTTGACCTTGTGACCATCTACAAAGTGTCTTCACT 2861
2461 GCGGTTTGTGGTCTGTACTCTCTGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
2862 AGGCTCATATGTTGTTTACAACTTTATCAACAGAGCGCGGACATGCAAGTGGG 2921
2521 AGGCTCATATGTTGTTTACAACTTTATCAACAGAGCGCGGACATGCAAGTGGG 2580
2922 GTCCCCCCTCAACGTTTCGGGAGCGCGGATGCCATCATCTCTCTCAAGTGTGCGGTT 2981

2581 ATCCCCCCTCAACGCTCGGGAGGCGCGATGCCATCATCTCTCTCATGTGGCGATC 2640
2982 CATCCAGAGTTAAATTTTGAATCAGCAAACTCTCTGCTCGCATACTCGGCCGCTCATG 3041
2641 CATCCAGAGTTAAATTTTGAATCAGCAAACTCTCTCTGCTCGCATACTCGGCCGCTCATG 2700
3042 GTGCTCCAGGCTGGCATAAACAGAGTGCCTGCTTCTGTCGCGCTCAAGGGCTCATCTG 3101
2701 GTGCTCCAGGCTGGCATAAACAGAGTGCCTGCTTCTGTCGCGCTCAAGGGCTCATCTG 2760
3102 GCATGATGTTAGTGCAGAAAGTGCCTGCGCGCTCATTTATGTCCAAATGCTTTCATGAAG 3161
2761 GCATGATGTTAGTGCAGAAAGTGCCTGCGGCTCATTTATGTCCAAATGCTTTCATGAAG 2820
3162 CTGGCGCGCTGACAGGTACGTACGTTTATAAACCATCTTACCCCACTGCGGGAATGSGCC 3221
2821 CTGGCGCGCTGACAGGTACGTACGTTTATAAACCATCTTACCCCACTGCGGGAATGSGCA 2880
3222 CAOCGCGCTTACAGAGCTTTCGCGTGGCTGAGTGCCTGCTGCTTCTGCGCATGAG 3281
2881 CGCGCGCGCTTACAGAGCTTTCGCGTGGCTGAGTGCCTGCTGCTTCTGCGCATGAG 2940
3282 ACCAGGTTCATCACTGGGAGCAGACACCGCTGGTGGGACATCATCTTGGGCTA 3341
2941 ACCAGGTTCATCACTGGGAGCAGACACCGCTGGTGGGACATCATCTTGGGCTG 3000
3342 CCGCTCTCGGCCGAAAGGAGATATTTTGGGACCGGCTGATGCTCGAAGG 3401
3001 CCGCTCTCGGCCGAAAGGAGATATCTCTGGGCGCGGCGGATAGTCTTGAAGG 3060
3402 CAAGGTTGGGACTCTCTGCGCGCTTACCGGCTTACTCCCAACAAACGCGGGGCTACTT 3461
3061 CCGGCTTGGGACTCTCTGCGCGCTTACCGGCTTACTCCCAACAGACGCGGGCTACTT 3120
3462 GGTTCATCATCACTAGCTTACAGCGCGGACAGACAGGCTGAGGAGGAGTTCAG 3521
3121 GGTTCATCATCACTAGCTTACAGCGCGGACAGACAGGCTGAGGAGGAGTTCAG 3180
3522 GTGGTTTTCACCGCAACAACTCTTCTGCGGCTGAGGCTGAGGCTGCTGCTG 3581
3181 GTGGTTTTCACCGGACAACTCTTCTGCGGCTGAGGCTGAGGCTGCTGCTG 3240
3582 GTCTACATGCGGCTGGCTGAGACCGCTAGCGGCTCAAAAGGCTCAATCACCAAGT 3641
3241 GTTTACCATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3300
3642 TACACCAATGTAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3701
3301 TACACTAATGTGACAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3360
3702 ACACCATGAGCTGTGGAGCTCGGACCTTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 3761
3361 ACACCATGAGCTGTGGAGCTCGGACCTTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 3420
3762 CCGTGGCGCGGAGGAGCAGACAGGAGGAGTCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3821
3421 CCGTGGCGCGGAGGAGCAGACAGGAGGAGTCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3480
3822 CTGAAAGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3881
3481 TTGAAGGCTCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3540
3882 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3941
3541 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3600
3942 ATGGAATCTACATGCGGCTTCTGAGGAGTCTTACAGACAACTCAACCCCGCGGCTGACCG 4001
3601 ATGGAATCTACATGCGGCTTCTGAGGAGTCTTACAGACAACTCAACCCCGCGGCTGACCG 3660
4002 CAGACATTCAGGAGGAGTCTGACCGCTCTTCTGAGGAGTCTTACAGACAACTCAACCCCGCGG 4061
3661 CAGTCACTTCAAGTGGGACCTTACCGCTCTTCTGAGGAGTCTTACAGACAACTCAACCCCGG 3720

Db 1 ATGAGCAGAAATCTTAAACCTCAAAGAAAAACCAAAAGCTAAACACCAAAACCGCGCCGACAG 60
Qy 402 GACGTCAGCTCCCGGCGGTGTCAGATCGTTGTGGAGTTTACCTGTTCGCGCGCAGG 461
Db 61 GACGTCAGCTCCCGGCGGTGTCAGATCGTTGTGGAGTTTACCTGTTCGCGCGCAGG 120
Qy 462 GCGCCAGGTTGGGTGTGCGCGCGACTAGGAAGGCTTCCGAGCGGTGCGAACTCGTGGA 521
Db 121 GCGCCAGGTTGGGTGTGCGCGCGCCAGGAAGACTTCCGAGCGGTGCGAACTCGTGGA 180
Qy 522 AGCGCAGACCTATCCCAAGGCTGCGCAACCGGAGGCGAGGCGCTGGGCTCAGCCCGGG 581
Db 181 AGCGCAGACCTATCCCAAGGCTGCGCGCGCGGAGGCGAGGCTGGGCTCAGCCCGGG 240
Qy 582 TACCCCTGCGCCCTCTATGGCAATGAGGCGCTTGGGTGGCGAGGATGGCTCTGTCAACC 641
Db 241 TACCCCTGCGCCCTCTATGGCAATGAGGCTTGGGTGGCGAGGATGGCTCTGTCAACC 300
Qy 642 CGCGGCTCCCGGCTAGTTGGGCGCCACGAGACCCCGCGTAGGTGCGGTAACTTGGGT 701
Db 301 CGCGGCTCCCGGCTAGTTGGGCGCCACGAGACCCCGCGTAGGTGCGGTAACTTGGGT 360
Qy 702 AAGGTCATCGATACCTTACATCGGCTTCGCGGATCTCATGGGTACATTCGCTCGTC 761
Db 361 AAGGTCATCGATACCTTACATCGGCTTCGCGGATCTCATGGGTACATTCGCTCGTC 420
Qy 762 GCGCGCCCTTAGGGGCGCTGCGAGGCGCTTGGCACAGGCTGTCGCGGTTCTGGAGGAC 821
Db 421 GCGCGCCCTTAGGGGCGCTGCGAGGCGCTTGGCACAGGCTGTCGCGGTTCTGGAGGAC 480
Qy 822 GCGGTGAATATGCAACAGGGAACATGCGCGGTGCTCTTCTATCTTCTCTCTTCTCT 881
Db 481 GCGGTGAATATGCAACAGGGAATGCGCGGTGCTCTTCTATCTTCTCTCTTCTCT 540
Qy 882 CTGCTGCTCTGTTGACATCCAGCTTCGCTTATGAAGTGCGCAACGTTCCCGGATA 941
Db 541 CTGCTGCTCTGTTGACATCCAGCTTCGCTTATGAAGTGCGCAACGTTCCCGGATA 600
Qy 942 TACCATGTACAGAAAGACGTGCTCAACTCAAGCATGTGTTATGAGGACGCGACGTGATC 1001
Db 601 TATCATGTACAGAAAGACGTGCTCAACGCAAGCATGTGTTATGAGGACGCGACGTGATC 560
Qy 1002 ATGCAATACCCCGGTGCTGCTGTTGTCAGGAGGTTAAAGCTCCCGTTGCTGGGTA 1061
Db 661 ATGCAATACCCCGGTGCTGCTGTTGTCAGGAGGTTAAAGCTCCCGTTGCTGGGTA 720
Qy 1062 GCGCTCACTCCAGCGTCGCGGCGAGGATGCGAGGTCGCCACTACGCAATACGAGCG 1121
Db 721 GCGCTCACTCCAGCGTCGCGGCGAGGATGCGAGGTCGCCACTACGCAATACGAGCG 780
Qy 1122 CACGTCGACTTGTGTTGGGACGCGTGTCTTCTGCTCGCTATGTACGTGGGGATCTC 1181
Db 781 CACGTCGACTTGTGTTGGGAGCGTGTCTTCTGTTCCGCTATGTACGTGGGGACCTC 840
Qy 1182 TCGGATCTATTTCTCTCTCCAGCTGTTCACTTCTGCTGCGCGGATAGACA 1241
Db 841 TCGGATCTATTTCTCTCTCCAGCTGTTCACTTCTGCTGCGCGGATAGACA 900
Qy 1242 GTGAGGACTGCACTGCTCAATCTATCCGCGCATGTATCAGGTCACCGATGCTTGG 1301
Db 901 TTACGAGACTGAATGCTCAATTTATCCCGCGCATGTGTCGGGTACCGTATGCTTGG 960
Qy 1302 GATATGATGATGAATGCTCACTCAACAGCGCTAGTGGGTGCGAGTTGCTCGGATC 1361
Db 961 GATATGATGATGAATGCTGCGCCACACAGCGCTAGTGGGTGCGAGTTATCTCGGATC 1020
Qy 1362 CACAGCTGTGCTGAGCATGTGGCGGGGCCACTGGGAGTCTTGGCGGCGCTTGGC 1421
Db 1021 CACAGCGCTGCTGAGCATGTGGCGGGGCCACTGGGAGTCTTGGCGGCGCTTGGC 1080
Qy 1422 TACTATTCCATGGTAGGGAAGTGGGCTAAGTTCTGATTTGGCGCTACTCTTGGCGGC 1481
Db 1081 TACTATTCCATGGCGGGGAAGTGGGCTAAGTTCTGATTTGATGCTACTTTTGTGTCG 1140

Qy 1482 GTTACGCGGAGAGCCACACAGAGCGGAGGTTGGCCGCCACACCACTTCGGGTTACG 1541
Db 1141 GTTACGCGGAGATCCACCTGACAGGCGGCGGAGCCAAAACCAACAGGCTCGTG 1200
Qy 1542 TCCCTTTTCTCATCTGGGCGTCTCAGAAAATCCAGCTTGTGAATACCAAGCGCAGCTGG 1601
Db 1201 TCCATGTTGCAAGTGGGCGCTCTCAGAAAATCCAGCTTATAAACACCAATGGGAGTTGG 1260
Qy 1602 CACATCACAGGACTGCGCTTAANTTCAATGACTCCCTCCAACTGGGTTCTTTGCGCG 1661
Db 1261 CACATCACAGGACTGCGCTGAATGCAATGACTCTCTCCAGACTGGGTTCTTTGCGCG 1320
Qy 1662 CTGTTTAAAGCACAAAGTTCAACTGCTCCGGGTGCGCGAGCGCATGCGGCTGCGCG 1721
Db 1321 CTGTTCTACACATAGTTTCACTCTGTCGGGTGCGCAGAGCGCATGCGGCTGCGCG 1380
Qy 1722 CCCATGACTGTTGCGCCAGGCGTGGGCGCCCATCACTATATCTAAGCCTAACAGCTCG 1781
Db 1381 ACCATTGACAAAGTTGCGACAGGATGGGTGCCATTAATTATGCTGAGCTAGCAGATCA 1440
Qy 1782 GATCAGAGGCTTATTTGCTGGCATTAACGGCTCGACCGTGTGGTGGTGGTACCGCGTCG 1841
Db 1441 GACCAGAGGCGCATATTTGCTGCACTTACCCACTCCCAATGTACCATCTGCTCGCTCG 1500
Qy 1842 CAGGTGTGTTGCTCAAGTGTATTTTACCCCAAGCCCTGTTGTGGTGGGACCAACCGAT 1901
Db 1501 GAGGTGTGGGCGGAGTGTACTGTTTCAACCGCTGCTGCTGGGAGACACCGAT 1560
Qy 1902 GCTTCCGCTGCTTACCTATAGCTTAGCTGGGGAGAAATGACAGAGCTGATGCTCTCAAC 1961
Db 1561 GCTTCCGCTGCTTACCTATAGCTTAGCTGGGGAGAAACGAGACTGAGCTGCTGCTCAAC 1620
Qy 1962 AACAGCGCTCGCGCACAGGCGAACTGTTGCGCTGTACATGGATCAATAGTACTGCTTC 2021
Db 1621 AACAGCGCGCGCGCAAGGCAACTGTTGCGCTGCAATGGATGAATAGCACCGGCTTC 1680
Qy 2022 ACTAAGAGCTGCGGAGGTCCCGCTGTAACTGCGGGGGGTGGTAAACCGACCTTGATC 2081
Db 1681 ACCAGACATGTGGGGGCGCGCTGTGTAACATCGGGGGGGTGGGACCAACACCTGACC 1740
Qy 2082 TGCCCCACGACTGCTTCCGGAAGCACCCGAGGCTACTTACACAAAATGTGGTTCGGG 2141
Db 1741 TGCCCCACGACTGCTTCCGGAAGCACCCGAGGCTACTTACACAAAATGTGGTTCGGG 1800
Qy 2142 CCCTGTTGACACCTAGTGTGCTAGTACACTACCAATACAGGCTTTGGCACTTACCCCTGC 2201
Db 1801 CCTTGGCTGACACCTAGTGTGCTAGTACACTACCAATACAGGCTTTGGCACTTACCCCTGC 1860
Qy 2202 ACTCTCAATTTTCCACTCTTAAAGGTTAGGATGTATGTGGGGGGGTGGAGCACAGGCTC 2261
Db 1861 ACTGTTACTTTTCCACTCTTAAAGGTTAGGATGTATGTGGGGGGGTGGAGCACAGGCTC 1920
Qy 2262 AATGCGCATGCAATGACTGCGAGGAGCGCTGTAACTTGGAGGAACAGGATAGGTCA 2321
Db 1921 AATGCTGATGCAATGAGGACCGGAGAGCGCTTGTGACTTGGAGGACAGGATAGGCG 1980
Qy 2322 GAAGTCAGCGCTGCTGCTGTCTACACAGAGTGGCAGATACTGCCCTGTGCTTCAAC 2381
Db 1981 GAGCTCAGCGCTGCTGCTGTCTACACAGAGTGGCAGGACTGCCCTGTTCCTTCAAC 2040
Qy 2382 ACCCTACCGGCTTTATCCACTGGTTTGAATCCATCTCCATCAGAAATCGTGGACGTCAA 2441
Db 2041 ACCCTACCGACTCTGCTCACTGGCTTGAATCCACTCCATCAGAAATCGTGGACGTCAA 2100
Qy 2442 TACCTGTAAGGTGAGGTGCTAGGTTGCTTCTCTTGGATCAAAATGGGAGTACATCTG 2501
Db 2101 TACCTTAAGGTGAGGTGCTAGGTTGCTTCTCTTGGATCAAAATGGGAGTATGCTGCTG 2160
Qy 2502 TTGCTTTTCTTCTCTGCGAGACGCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2561
Db 2161 TTGCTTTTCTTCTCTGAGCGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220

QY 2562 ATAGCCAGGCTGAGGCGGCTTAGAGAACTTGGTGTCTCAATGCGGCGTCCGTGGCC 2621
DB 2221 ATAGCCAGGCGGAGCGGCTTGGAGAACTTGGTGTCTCAATTGGGCGTCTGGGCC 2280
QY 2622 GGAGCGCATGTATCTCTCTTCTTGTGTCTTCTGGCGCGCTCGGTACATTAAGGGC 2681
DB 2281 GGCGCATGTCATCTCTCTCTTGTGTCTTCTGTGGCGCTGTGTACATCAAGGC 2340
QY 2682 AGGCTGCTCTGCGGCGGCTATGCTTTTATGCGGTATGCGCGTCTCTCTCTCTA 2741
DB 2341 AGGCTGCTCTGCGGCGGCTATGCTTTTATGCGGTATGCGCGTCTCTCTCTCTA 2400
QY 2742 CTGCGCTTACCAACAGAGCTTACGCGCTTGGACCGGAGATGGTGTATCGTGGCGGGT 2801
DB 2401 CTGCGCTTACCAACAGAGCTTACGCGCTTGGACCGGAGATGGTGTATCGTGGCGAGC 2460
QY 2802 GCGGCTCTGTAGTCTGTATCTTGTACCTTGTACCATACATACATACATACATACAT 2861
DB 2461 GCGGCTCTGTAGTCTGTATCTTGTACCTTGTACCATACATACATACATACATACAT 2520
QY 2862 AGGCTCATATGCTGTACATATCTTTATACACAGAGCGGAGCGGCATGCAAGTGTGG 2921
DB 2521 AGGCTCATATGCTGTACATATCTTTATACACAGAGCGGAGCGGCATGCAAGTGTGG 2580
QY 2922 GTCCCGCCCTCAAGCTTGGGAGGCGCGAGTGCATCATCTCTCAAGTGTGGCGGT 2981
DB 2581 ATCCCGCCCTCAAGCTTGGGAGGCGCGAGTGCATCATCTCTCAAGTGTGGCGAGTC 2640
QY 2982 CATCCAGAGTAAATTTTGTACATCACCAGAGTCCGTCTGCGCTCAAGGCTCATCTCGT 3041
DB 2641 CATCCAGAGTAAATTTTGTACATCACCAGAGTCCGTCTGCGCTCAAGGCTCATCTCGT 2700
QY 3042 GTGCTCCAGGCTGGCATAACAGAGTCCGTCTGCGCTCAAGGCTCATCTCGT 3101
DB 2701 GTGCTCCAGGCTGGCATAACAGAGTCCGTCTGCGCTCAAGGCTCATCTCGT 2760
QY 3102 GCATGCTATGTAGTGGAGAAAGTCCGCGGGGTCAATATGTCGCGCTCATGAG 3161
DB 2761 GCATGCTATGTAGTGGAGAAAGTCCGCGGGGTCAATATGTCGCGCTCATGAG 2820
QY 3162 CTGGGCGGCTGACAGTACCTTATATACCATCTTACCCACATCGCGGAGTGGGCC 3221
DB 2821 CTGGGCGGCTGACAGTACCTTATATACCATCTTACCCACATCGCGGAGTGGGCC 2880
QY 3222 CACGCGGCGCTACGAGACCTTGGCGGTGAGCGCGGTCTCTCTCGCGCATGGAG 3281
DB 2881 CCGCGGCGCTACGAGACCTTGGCGGTGAGCGCGGTCTCTCTCGCGCATGGAG 2940
QY 3282 ACCAAGTCTACCTGGGAGCAGACACCGCTGGTGTGGGACATCATCTGGGTCTA 3341
DB 2941 ACCAAGTCTACCTGGGAGCAGACACCGCGCGGTGTGGGACATCATCTGGGTCTG 3000
QY 3342 CCGCTCTCGCGCGAAGGGAAGAGATATTTTGGGACCGGTGATGTCTCGAAGGG 3401
DB 3001 CCGCTCTCGCGCGAAGGGAAGAGATATCTCTGGCGCGCGGTATGTCTTGAAGGG 3060
QY 3402 CAAAGGTGGGACCTCTTGGCGCATACCGCTCTCTCCCAACAAACCGCGCGGTACTT 3461
DB 3061 CCGGCGGTGGGACCTCTTGGCGCATACCGCTCTCTCCCAACAAACCGCGCGGTACTT 3120
QY 3462 GGTGTGATCATCTAGCTCAGCGCGGAGCAGACACCGGTGGAAGGAGGTTCAA 3521
DB 3121 GGTGTGATCATCTAGCTCAGCGCGGAGCAGACACCGGTGGAAGGAGGTTCAA 3180
QY 3522 GTGGTTTCTACCGCAACAATCTTCTCTGGCGACTTGCATCAACCGGCTGTGTGGACT 3581
DB 3181 GTGGTTTCTACCGCAACAATCTTCTCTGGCGACTTGCATCAACCGGCTGTGTGGACT 3240
QY 3582 GTCTACCATGGCGGTGGTCTGAAAGACCTTAGCGGTCCAAAGGTCCCAATCAACCAATG 3641
DB 3241 GTCTACCATGGTGGTCTGAAAGACCTTAGCGGTCCAAAGGTCCCAATCAACCAATG 3300
QY 3642 TACACCAATGTAGACCTGGAACCTCTGTGGTGGAGGCGCGCGCGCGGCTCTCATG 3701

DB 3301 TACATTAATGTGACACAGGACCTCGTGGCTGGCCCAAGCCCGCGCGGCTCTCTTG 3360
QY 3702 ACACATGACAGTGTGGAGCTCGGACCTTACTTGGTACGAGACATCTGTATGTCAAT 3761
DB 3361 ACACCATGACCTGTGGAGCTCAGACCTTACTTGGTACGAGACATCTGTATGTCAAT 3420
QY 3762 CCGGTCTGGCGGAGCGAGCAGCAGGGAAGTCTACTCTCCCGCAGGCGCGCTCTCTAC 3821
DB 3421 CCGGTCTGGCGGAGCGGAGCAGCAGTGGGAGGCTCTCTCTCCCGAGGCTGTCTCTAC 3480
QY 3822 CTGAAGGCTCTCTGGGTGTCTATCTTGGTCTTGGGCGACGTCTGTGGGCTCTTC 3881
DB 3481 TTGAAGGCTCTCTGGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3540
QY 3882 CCGGCTGTGTGTGACCCCGGGGGTGGAGGCGGTGGACTTCTATACCCGTGTAGTCT 3941
DB 3541 CCGGCTGTGTGTGACCCCGGGGGTGGAGGCGGTGGACTTCTATACCCGTGTAGTCT 3600
QY 3942 ATGGAACCTACCATGCGGTCTCTGGTCTTACAGACAACTCAACCCCGCGGTGTACCG 4001
DB 3601 ATGGAACCTACCATGCGGTCTCTGGTCTTACAGACAACTCAACCCCGCGGTGTACCG 3660
QY 4002 CAGACATTCGAAGTGGACATCTGCAAGCTCTCTCTGCGAGGCGGAGCAGCAGAGT 4061
DB 3661 CAGTCAATTCGAAGTGGACATCTGCAAGCTCTCTCTGCGAGGCGGAGAGTACTAAGT 3720
QY 4062 CCGGCTGTGTGTGACCCCGGGGGTGGAGGCGGTGGACTTCTATACCCGTGTAGTCT 4121
DB 3721 CCGGCTGTGTGTGACCCCGGGGGTGGAGGCGGTGGACTTCTATACCCGTGTAGTCT 3780
QY 4122 ACCTTAGGTTTGGGCGGTATATCTCAAGGACACAGGTATCGACCTAACATCAAGT 4181
DB 3781 ACCTTAGGTTTGGGCGGTATATCTCAAGGACACAGGTATCGACCTAACATCAAGT 3840
QY 4182 GGGTAAAGACCATTACCAAGGCGGTCTTACGTCTCTCTCTCTCTCTCTCTCTCTCT 4241
DB 3841 GGGTAAAGACCATTACCAAGGCGGTCTTACGTCTCTCTCTCTCTCTCTCTCTCTCT 3900
QY 4242 GCGAGCGTGGCTGTCTGGGCGGTATGATCATATATGATGATGATGATGATGATGAT 4301
DB 3901 GCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3960
QY 4302 ACTGACTCGATACCATCTTGGGCTCTGCGACAGTCTCTGAGCAAGCGGAGTGGGA 4361
DB 3961 ACTGACTCGATACCATCTTGGGCTCTGCGACAGTCTCTGAGCAAGCGGAGTGGGA 4020
QY 4362 GCGGCGCT 4421
DB 4021 GCGGCGCT 4080
QY 4422 AATATCGAGGAATAGGCTGTCTCAACATGAGAGATCCCTTCTATGAGCAAGGCTATC 4481
DB 4081 AATATCGAGGAATAGGCTGTCTCAACATGAGAGATCCCTTCTATGAGCAAGGCTATC 4140
QY 4482 CCGATGAGGCGATCAAGGGGGGAGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4541
DB 4141 CCGATGAGGCGATCAAGGGGGGAGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4200
QY 4542 GAGGAGCTCCCGCAAGAGTGCAGAGGCTCGGAGTGAAGCTGTAGCATATTTACCGGGC 4601
DB 4201 GAGGAGCTCCCGCAAGAGTGCAGAGGCTCGGAGTGAAGCTGTAGCATATTTACCGGGC 4260
QY 4602 CTTGATGTCT 4661
DB 4261 CTTGATGTCT 4320
QY 4662 ATGACGGGTTTACCGGCGATTTTGTAGTCTGATGATGATGATGATGATGATGATGAT 4721
DB 4321 ATGACGGGTTTACCGGCGATTTTGTAGTCTGATGATGATGATGATGATGATGATGAT 4380
QY 4722 ACAGTGCATCTAGCTGTGATCTCCACTTCAACATGAGAGCAGGCGGCGGCGGCTCCAG 4781

4381 ACAGTCGACCTTACGCTTGGATCCACACTTACCAATTGAGACGACGACCGCTCCTCAAGAC 4440
4782 GCGGTGTGCGCTCGCAACGCGAGGTAGAACTGGCAGGAGTAGAGTGGATCTACAGG 4841
4441 GCACTGTGCGCTCGCAACGCGAGGTAGAACTGGCAGGAGTAGAGTGGATCTACAGG 4500
4842 TTTGTGACTCCAGGAGAACGCGCTCGGGCACTGTTCGATCTTCTCGTCTCTGTGAGTGC 4901
4501 TTTGTGACTCCGAGAGAACGCGCTCGGGCACTGTTCGATCTTCTCGTCTCTGTGAGTGC 4560
4902 TATGACGGGGCTGTGTGTGTGTATGAGCTACAGCCCGCTGAGACCTCGGTTAGTGTGCG 4961
4561 TATGACGGGGCTGTGTGTGTGTATGAGCTACAGCCCGCTGAGACCTCGGTTAGTGTGCG 4620
4962 GCTTACCTTAATACACACGAGGTGCGCTCTGCCAGGACATCTGGAGTTCCTGGAGAGC 5021
4621 GCCTACCTGAAACACACAGGTTGCGCTTTGCCAGGACCACTGGAGTTCCTGGAGAGT 4680
5022 GTCTTTCAGGCGCTCACCCACATAGATGCCCACTTCTGTCTCCAGACTAAACACAGCAGGA 5081
4681 GTCTTTCAGGCGCTCACCCACATAGATGCCCACTTCTGTCTCCAGACTAAACACAGCAGGA 4740
5082 GACAACTTCTCTACCTGGTGGCATATCAAGCTACAGTGTGCGGCGAGGCTCAAGTCCA 5141
4741 GACAACTTCTCTACCTGGTGGCATATCAAGCTACAGTGTGCGGCGAGGCTCAAGTCCA 4800
5142 CCTCATCTGCGGACCAATGTGGAAGTGTCTCATACGGCTGAACTACACTGACAGG 5201
4801 CCTCATCTGCGGACCAATGTGGAAGTGTCTCATACGGCTGAACTACACTGACAGG 4860
5202 CCAACACCCCTGCTGTATAGGCTAGGACCGCTCCAAATGAGGTATCTCTCACACACCC 5261
4861 CCAACACCCCTGCTGTATAGGCTAGGACCGCTCCAAATGAGGTATCTCTCACACACCC 4920
5262 ATAACTAAATCATATGGAATGCGATGCGGTGACCTGAGGTGCTCACTAGCACCTGG 5321
4921 ATAACTAAATCATATGGAATGCGATGCGGTGACCTGAGGTGCTCACTAGCACCTGG 4980
5322 GTGCTGTAGGCGAGTCTTGCAGCTTTGCGCGCATCTGCTGAGTGTGCGGAGGAGTGTG 5381
4981 GTGCTGTAGGCGAGTCTTGCAGCTTTGCGCGCATCTGCTGAGTGTGCGGAGGAGTGTG 5040
5382 GTCAATGTGGGAGGATCATCTTGTCCGGGAAGCCAGCTGTGCTTCCCGACAGGGAATC 5441
5041 GTCAATGTGGGAGGATCATCTTGTCCGGGAAGCCAGCTGTGCTTCCCGACAGGGAATC 5100
5442 CTCTACAGGAGTGTGATGAGAGGAAAGTGTGCTTCACTCACTTCTTCAATGAGAGAG 5501
5101 CTCTACAGGAGTGTGATGAGAGGAAAGTGTGCTTCACTCACTTCTTCAATGAGAGAG 5160
5502 GGAATGCAAGCTCGCGAGCAATTAAGCAAAAGGCGCTCGGCTGTGCTCAAAAGGCGCAC 5561
5161 GGAATGCAAGCTCGCGAGCAATTAAGCAAAAGGCGCTCGGCTGTGCTCAAAAGGCGCAC 5220
5562 AAGCAAGCGAGGCTGTGCTTCCCTGTGTGAGTCCAAAGTGGCGAGGCGCTTGAACCTTC 5621
5221 AAAAAGCGAGGCTGTGCTTCCCTGTGTGAGTCCAAAGTGGCGAGGCGCTTGAACATTC 5280
5622 TGGGGAGAGCAATGTGAAATTTATATGAGGGAATACAGTACCTAGAGGCTTATCCACT 5681
5281 TGGGGAGAGCAATGTGAAATTTATATGAGGGAATACAGTACCTAGAGGCTTATCCACT 5340
5682 CTGCTGAAACCCCGCATAGCATCATTTGATGGCATTTACAGCTTCTATCATAGCCCG 5741
5341 CTGCTGGAACCCCGCATAGCATCATTTGATGGCATTTACAGCTTCTATCACAGCCCG 5400
5742 CTCACACCCAAACACCCCTCTCTTTTAACTATCTTTGGGGAGTAGGCTGGCTGCCAACTC 5801
5401 CTCACACCCAAAGTACCTCTCTTTTAACTATCTTTGGGGAGTAGGCTGGCTGCCAACTC 5460
5802 GCTCTCCAGGCTGTGCTGAGCTTGTGCGGCGCGGAGAGGCTGTGCTTTAGGAGC 5861
5461 GCGCGCGCGCTTGTGCGCTTGTGCGGCGCGGAGTGTGCGGCGCGGCTGTGCGGCTTGTGCG 5520

5862 AGCATGGCTTCGGAGAGTGTCTCTGGGACATCTTGGCGGCTATGSGGCGAGGAGTACCC 5921
5521 AGCATGGCTTCGGAGAGTGTCTCTGGGACATCTTGGCGGCTATGAGGAGGAGTACCC 5580
5922 GGGGCACTCGTGGCTTTTAAAGTTCATGAGCGGAGGTGCGCTCCACCGAGGAGCTGCT 5981
5581 GGGGCGCTCGTGGCTTTTAAAGTTCATGAGCGGAGATGCCCTCCACCGAGGAGCTGCT 5640
5982 AACTTACTCTCGCATCTCTCTCTGGTGGCTTGGTCTGGGGTCTGTGCGGAGCA 6041
5641 AACTTACTCTCGCATCTCTCTCTGGGCGCTTGGTCTGGGGTCTGTGCGGAGCA 5700
6042 ATACTGCTCGGACAGTGGGCGCGGAGAGAGGGGCTGTGCAAGTGGATGAACCGCTGATA 6101
5701 ATACTGCTCGACACAGTGGGTCGGGAGAGGGGCTGTGCAAGTGGATGAACCGCTGATA 5760
6102 GCGTTCCTTTGGGGGTAAACAGCTCTCCCTACGCACTATGTGCTGAGAGGAGCT 6161
5761 GCGTTCCTTTGGGGGTAAACAGCTCTCCCTACGCACTATGTGCTGAGAGGAGCT 5820
6162 GCAGCAGTGTCTAGATCTCTCTAGCTTACCATCTCACTCACTCACTCACTCACTCACT 6221
5821 GCAGCAGTGTCTAGATCTCTCTAGCTTACCATCTCACTCACTCACTCACTCACTCACT 5880
6222 CACAGTGGATTAATGAGGACTGCTCTACCGCATGCTCCGCTCGTGGCTAAGGAGTGT 6281
5881 CACAGTGGATTAATGAGGACTGCTCTACCGCATGCTCCGCTCGTGGCTAAGGAGTGT 5940
6282 TGGATTTGATATGACAGGTTTGAATGATTTCAAGACCTGGGTCAGTCCAACTCTCT 6341
5941 TGGGACTGGATATGACAGGTTTGAATGATTTCAAGACCTGGGTCAGTCCAACTCTCT 6000
6342 CCGCGGTTACGGGAGTCCCTTCTCTGTCATGCCAAGCGGCTACAGGAGTCTGCGG 6401
6001 CCGCAGCTACTGGAGTCCCTTCTCTGTCGACACGCGGTCACAGGAGTCTGCGG 6060
6402 GGGGACGCGATCATGCAAAACACCTGCCATCGGAGCACAGATCGCGGACATGTCAAA 6461
6061 GGAGACGCGATCATGCAAAACACCTGCCATGAGGAGCACAGATCGCGGACATGTCAAA 6120
6462 AACGGTTCCATAGGATCGTAGGCTTAGAACTTGCGAGCAACAGTGGGCGAGAACTTC 6521
6121 AACGGTTCCATAGGATCGTAGGCTTAGAACTTGCGAGCAACAGTGGGTCAGAACATTC 6180
6522 CCCATCAACGATACACACGAGGACCTTGCACACCTCCCGGCGCCCACTATTCAGG 6581
6181 CCCATCAACGATACACACGAGGCGCCCTGCACACCTCTCCAGCGCCAACTATTCAGG 6240
6582 GCGCTATGGCGGCTGCTGTGAGGAGTACGTGGAGGTACCGCTGTGGGGGATTTCCAC 6641
6241 GCGCTGTGGCGGCTGCGCTGAGGAGTACGTGGAGGTACCGCGGTGGGGGATTTCCAC 6300
6642 TACGTGACGGGATGACCACTGACAACTAAAGTGGCCATGCCAGGTTCCGCGCCCGAA 6701
6301 TACGTGACGGGATGACCACTGACAACTAAAGTGGCCATGCCAGGTTCCGCGCTCCGAA 6360
6702 TCTTTCAGGAGGTGATGAGTGGCTTGCACAGGTACCGTCCGCGTGCAGAACCTCTT 6761
6361 TCTTTCAGGAGGTGATGAGTGGCTTGCACAGGTACCGTCCGCGTGCAGAACCTCTT 6420
6762 CTAGCGAGGAGCTCACGTTCCAGTCCGGCTCAACCAATCTTGGTGGGTCGAGCTC 6821
6421 CTAGCGAGGAGGTACATTTCCAGTCCGGCTCAACCAATCTTGGTGGGTCGAGCTC 6480
6822 CCATGCGAGCGCAACCGAGGTAACTGCTTACTTCCATGCTCACTGCTCCCTCCAC 6881
6481 CCATGCGAGCGCAACCGAGGTAACTGCTTACTTCCATGCTCACTGCTCCCTCCAC 6540
6882 ATTTACGAGAGAGCGGTAACTGCTGCTGAGGAGTCTCCCTCTTTCAGGAGC 6941
6541 ATCAAGAGAGAGCGGTAACTGCTGCTGAGGAGTCTCCCTCTTTCAGGAGC 6600

QY 6942 TCATCAGCTAGCCAGTTGCTCGGCTTCTTTGAAGCGACATGCATCACCACCATGAC 7001
 Db 6901 TCTTCAGCTAGCCAGTTGCTCGGCTTCTTTGAAGCGACATGCATCACCACCATGTC 5660
 QY 7002 TCCCCGAGCGCTGACCTCATCGAGGCAACCTCTTGTGGCGGACGAGATGGCGGAAAC 7061
 Db 6661 TCTTCGAGCGCTGACCTCATCGAGGCAACCTCTTGTGGCGGACGAGATGGCGGAAAC 6720
 QY 7062 ATCACTCGGCTGAGTACAGAGATAGAGTATGATTTCTGGACTTTTGGAAACGGCTTCAC 7121
 Db 6721 ATCAACCGGCTGGAGTCCGAGAAACAAGTGGTATCTCTGGACTTTTTCGACCGCTTCGA 6780
 QY 7122 GCGAGGGGATCAGAGGGAGATATCGCTCGCGCGGAGATCTCTGGAATAATCCAGGAAG 7181
 Db 6781 GCGAGGAGATCAGAGGGAGATATCGCTTCGCGGGAGATCTCTGCGGAATCCAGGAAG 6840
 QY 7182 TTTCCCTCAGCTTGCCATPATGGGACGCGCGGACTTACAAATCTCTCACTGCTAGAGTCC 7241
 Db 6841 TTTCCCGCAGCGATGCCATCTGGGCGCGCGGATTAACAACCTCCACTGTTAGAGTCC 6900
 QY 7242 TCGAAGGACCGGACTACGTCCTCGGCTGTACAGGATGCCATTTGCCACCTACCAAG 7301
 Db 6901 TCGAAGGACCGGACTACGTCCTCGGCTGTACAGGATGCCGTTTGGCACTTATCAAG 6960
 QY 7302 GCTTCCTCAATACCACTCCACGAGAGAAAGAGACGGTTCTCTTGAACAGAAATCCAAATGTG 7361
 Db 6961 GCCCTTCCATACCACTCCACGAGAGAAAGAGACGGTTCTCTTGAACAGAAATCCAAATGTG 7020
 QY 7362 TCTTCTGCCCTTGGCGAGCTCGGCACTAAGACCTTGGTATGCTCGGATCGTGGCGGTT 7421
 Db 7021 TCTTCTGCCCTTGGCGAGCTCGGCACTAAGACCTTGGCGAGCTCCGAATCATCGGCGGTC 7080
 QY 7422 GATAGCGGACCGGACCGCTCTCTGACTCGGCTCCGACGAGGTCACAAAGGATCC 7481
 Db 7081 GACAGCGGACCGGACCGGCTCTCTGACTCGGCTCCGACGAGGTCACAAAGGATCC 7140
 QY 7482 GACGTTGAGTCGTAATCTCTCATGCGGCTTGAAGGGAGCGGGGACCGGATCTC 7541
 Db 7141 GACGTTGAGTCGTAATCTCTCATGCGGCTTGAAGGGAGCGGGGACCGGATCTC 7200
 QY 7542 AGGACCGGCTTGGTCTACGGTGAAGTGAAGGCTAGTGAAGTATGCTGCTGCTCA 7601
 Db 7201 AGTGAACGGCTTGGTCTACGGTGAAGTGAAGGCTAGTGAAGTATGCTGCTGCTCA 7260
 QY 7602 ATGTCCTATAGTGAAGCGGCTGATCAAGCGGCTGCGTGGGAGGAAAGTAAAGCTG 7661
 Db 7261 ATGTCCTATACAGGACGCGCTTGTATGCGGCTGCGTGGGAGGAAAGTAAAGCTG 7320
 QY 7662 CCCATCAACCGGTTGAGCACTTTGCTGCGGACCATATATGCTTATGCGGACCAACA 7721
 Db 7321 CCCATCAACCGGTTGAGCACTTTGCTGCGGACCATATATGCTTATGCGGACCAACA 7380
 QY 7722 TCCCGAGCGGACGCTCCGCGAAGAGTCACTTTGACAGATTGCAAGTCTCTGAT 7781
 Db 7381 TCTCCGAGCGGCTGCGGCAAGAGTCACTTTGACAGATTGCAAGTCTCTGAT 7440
 QY 7782 GATCATTAACCGGAGCTACTCAAGGAGATGAAGGCGGCTCCACAGTTTAAGGCTAAG 7841
 Db 7441 GACCACTACCGGAGCTGCTCAAGGAGATGAAGGCGGCTCCACAGTTTAAGGCTAAG 7500
 QY 7842 CTTCTATCTATAGAGGAGCTGGAAGCTGAGCGGCGGACATTTGGGCCAATTCBAATTT 7901
 Db 7501 CTCCTATCCGTAAGAGAGCTGCAAGCTGACGCGGCGGACATTTGGGCCAATTCBAATTT 7560
 QY 7902 GGCTATGGGCAAGAGGCTCGGAGCTATCCAGCGGCGGCTTAAACCATCGCTCC 7961
 Db 7561 GGCTATGGGCAAGAGGCTCGGAGCTATCCAGCGGCGGCTTAAACCATCGCTCC 7620
 QY 7962 GTGTGGGAGGACTTCTCTGGAAGACACTGAAACACCAATTTGACACCAACCATCATGCGCAAAA 8021
 Db 7621 GTGTGGGAGGACTTCTCTGGAAGACACTGTAACCAATTTGACACCAACCATCATGCGCAAAA 7680
 QY 8022 AGTGAGGTTTCTGCGTCCAAACAGAGAGGAGCGGCGGAGCTCGGCTTATCGTA 8081

Db 7681 AATGAGGTTTCTGTTCTCAACACAGAGAAAGAGGCGCTTAAGCCAGCGCGCTTATCGTA 7740
 QY 8082 TTCCACAGCTGGGAGTTCTGTTATGCGAGAGATGGCCCTTTACGAGTGTCTTCCACC 8141
 Db 7741 TTCCACAGATCTGGGAGTCCGTTATGCGAGAGATGGCCCTTTATGATGTGTCTTCCACC 7800
 QY 8142 CTTCTCAGGCGCTGATGGGCTCTCTATACGATTTCAATATCTCCCAAGCAGCGGCTC 8201
 Db 7801 CTTCTCAGGCTGATGGGCTCTCTATACGATTTCAATATCTCTCTGGGACGCGATC 7860
 QY 8202 GAGTTCCTGGTGAATACCTGGAATCAAGAAATGCGCTATGGGCTTCTCATATGACACC 8261
 Db 7861 GAGTTCCTGGTGAATACCTGGAATCAAGAAATGCGCTATGGGCTTCTCATATGACACT 7920
 QY 8262 CGCTGTTTGAATCAACGCTCACTCAGAGTGAATTCGTTGTTGAGGAGTCAATTTACCAA 8321
 Db 7921 CGCTGTTTGAATCAACGCTCACTCAGAGTGAATTCGTTGTTGAGGAGTCAATTTACCAA 7980
 QY 8322 TGTGTGACTTGGCGCGGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8381
 Db 7981 TGTGTGACTTGGCGCGGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8040
 QY 8382 ATCGGGGCTCCCTGACTTACTCAAAAGGAGAGTGGGCTTATCGCGGTCGCGGCA 8441
 Db 8041 ATCGGGGCTCCCTGACTTACTCAAAAGGAGAGTGGGCTTATCGCGGTCGCGGCG 8100
 QY 8442 AGTGGGCTGCTGACCACTAGCTGCGGTAATACCTCACAATGTTTGAAGGCGCACTGCA 8501
 Db 8101 AGCGGGCTGCTGACCACTAGCTGCGGTAATACCTCACAATGTTTGAAGGCGCACTGCA 8160
 QY 8502 GCCTGTCGAGCTGCAAGCTCCAGGACTGCAAGTCTCGTGAACGAGAGACGACTTGTTC 8561
 Db 8161 GCCTGTCGAGCTGCAAGCTCCAGGACTGCAAGTCTCGTGAACGAGAGACGACTTGTTC 8220
 QY 8562 GTTATCTGTAAGCGCGGAGACCGAGAGGAGTGGCGGCGCTTACAGGCTTCAAGGAG 8621
 Db 8221 GTTATCTGTAAGCGCGGAGACCGAGAGGAGTGGCGGCGCTTACAGGCTTCAAGGAG 8280
 QY 8622 GCTATGACTAGTATTTCCGCGCGGAGTCCGCGGCGCTTACAGGCTTCAAGGAGTGGAG 8681
 Db 8281 GCTATGACTAGTATTTCCGCGCGGAGTCCGCGGCGCTTACAGGCTTCAAGGAGTGGAG 8340
 QY 8682 CTGATTAATCATGTTCTTCCATGTTCTAGTGGCGGACGATGCTTGGCAAGAGGTA 8741
 Db 8341 CTGATTAATCATGTTCTTCCATGTTCTAGTGGCGGACGATGCTTGGCAAGAGGTA 8400
 QY 8742 TACTACTCAGCGTGACCCACCGGCTTGGCGGCGCTTGGCGGAGACAGCTAGA 8801
 Db 8401 TACTACTCAGCGTGATCCACCGGCTTGGCGGCGCTTGGCGGAGACAGCTAGA 8460
 QY 8802 CACACTCCATCAACTTTGGCTAGGCAATATCATCATGATGCGGCGGCGGCTTGGCGGA 8861
 Db 8461 CACACTCCATCAACTTTGGCTAGGCAATATCATCATGATGCGGCGGCGGCTTGGCGGA 8520
 QY 8862 AGGATGATTCGATGACTGACTTTTCTCCATGCTTCTAGCTCAAGAGGAGTGGGAAA 8921
 Db 8521 AGGATGATTCGATGACTGACTTTTCTCCATGCTTCTAGCTCAAGAGGAGTGGGAAA 8580
 QY 8922 GCCTGATTTGTCAGATCTACGGGCTTGTACTCCATTGAGGCGCTTACCTTACCTCAG 8981
 Db 8581 GCCTGATTTGTCAGATCTACGGGCTTGTACTCCATTGAGGCGCTTACCTTACCTCAG 8640
 QY 8982 ATCATTTGAAGGACTTCCATGCTTATGCGGCTTATGCTTCCAGGTTTCTTCCAGGTTAG 9041
 Db 8641 ATCATTTGAAGGACTTCCATGCTTATGCGGCTTATGCTTCCAGGTTTCTTCCAGGTTAG 8700
 QY 9042 ATCAATAGGTTGGCTTATGCTTCCAGGAACTTGGGGTACCACTTGGCGAACTTGGAGA 9101
 Db 8701 ATCAATAGGTTGGCTTATGCTTCCAGGAACTTGGGGTACCACTTGGCGAACTTGGAGA 8760
 QY 9102 CATCGGCGGAGAGTGTCCGCTTAACTTGTCCCGGCGGAGGCGGCGGAGCTTGTGT 9161

Db 8761 CATCGGCGCAGAGCGTCCGCGCTAGGCTACTGTCCAGGAGGAGGCGCCGCACTGT 8820
Qy 9162 GCGAGATACCTTTTAACTGGGAGCTAAGGACCAAGCTTAACTCAGTCCGATCCGCGCC 9221
Db 8821 GCGAAATACCTTTTCAACTGGGAGTAAACCAAACTTAAACTCAGTCCGCGCT 8880
Qy 9222 CCGTCCGAGCTGGACTTCTGTGGCTGGTTCGTCGCTGGTTCAGCGGGGAGACATAT 9281
Db 8881 GCGTCCGCGCTGGACTTCTGTGGCTGGTTCGTCGCTGGTTCAGCGGGGAGACATAT 8940
Qy 9282 CACAGCTGTCTCGTCCGCGAGCCCGCTGGTTCGTCGCTGGTTCAGTCTTCTCTGTA 9341
Db 8941 CACAGCTGTCTCGTCCGCGAGCCCGCTGGTTCAGTCTGTGCTACTCTACTTCTGTA 9000
Qy 9342 GGGGTAGCATTTACTGCTCCCAACCGA 9371
Db 9001 GGGGTAGCATTTACTGCTCCCAACCGA 9030

RESULT 15

US-09-315-850-13
; Sequence 13, Application US/09315850
; Patent No. 6217872
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLland &
; ADDRESS: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,850
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,686
; FILING DATE: 01-AUG-1997
; APPLICATION NUMBER: US/08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McLeLland, Le-Nhung

; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 900703G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9030 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from genomic RNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..9030
; OTHER INFORMATION: /note: "sequence = 333 - 9362 of
; OTHER INFORMATION: SEQ ID NO: 1"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1...9030
US-09-315-850-13
Query Match 81.4%; Score 7809.2; DB 3; Length 9030;
Best Local Similarity 91.6%; Pred. No. 0;
Matches 8267; Conservative 0; Mismatches 763; Indels 0; Gaps 0;
Qy 342 ATGAGCAGGATCTCTAAACCTCAAGAGAAACCAACCTACACCAACCGCGCCACAG 401
Db 1 ATGAGCAGGATCTCTAAACCTCAAGAGAAACCAACCTACACCAACCGCGCCACAG 60
Qy 402 GACGTCAGATTTCCGCGCGGTGGTTCAGATCGTTGGTGGATTTACCTGTCGCGCGAGG 461
Db 61 GACGTCAGATTTCCGCGCGGTGGTTCAGATCGTTGGTGGATTTACCTGTCGCGCGAGG 120
Qy 462 GCGCCGAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 521
Db 121 GCGCCGAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
Qy 522 AGGCGACAACTATCCCAAGGCTCGCGAGCCGAGGCGAGGCGCTGGGCTCAGCCCGGG 581
Db 181 AGGCGACAACTATCCCAAGGCTCGCGAGGCTCGCGAGGCGAGGCGCTGGGCTCAGCCCGGG 240
Qy 582 TACCTTTGGCCCTCTATGGCAATGAGGCGCTGGGCGTGGGCGAGGATGGTCTCTGTCACCC 641
Db 241 TACCTTTGGCCCTCTATGGCAATGAGGCGCTGGGCGTGGGCGAGGATGGTCTCTGTCACCC 300
Qy 642 CCGGCTCCGCGCTAGTGGGCGCCACGAGACCCCGCGGTAGTTCGCTAACTGGGT 701
Db 301 CCGGCTCCGCGCTAGTGGGCGCCACGAGACCCCGCGGTAGTTCGCTAACTGGGT 360
Qy 702 AAGGTCATCATACCTTTACATGCGGCTTCGCGGATCTCATGGGCTACATTCGCTCGTC 761
Db 361 AAGGTCATCATACCTTTACATGCGGCTTCGCGGATCTCATGGGCTACATTCGCTCGTC 420
Qy 762 GCGCGCCCTTAGGGGGCGCTGCCAGGCGCTTGGCAGACGCTTCCGGGTCTGGAGGAC 821
Db 421 GCGCGCCCTTAGGGGGCGCTGCCAGGCGCTTGGCAGACGCTTCCGGGTCTGGAGGAC 480
Qy 822 GCGGTGAATATGCAACAGGGAACCTTGGCCGGTTCCTTCTCTATCTTCTCTTGGCT 881
Db 481 GCGGTGAATATGCAACAGGGAACCTTGGCCGGTTCCTTCTCTATCTTCTCTTGGCT 540
Qy 882 CTGCTGTCTGTTTGACCATCCAGCTTCGCTTATGAAGTGGCAGGAGTTCGCGGATA 941
Db 541 CTGCTGTCTGTTTGACCATCCAGCTTCGCTTATGAAGTGGCAGGAGTTCGCGGATA 600
Qy 942 TACCATGTACGAGACGACTGTCCAACTCAAGCATTTGTGTATGAGGCGAGCGAGTATC 1001
Db 601 TATCATGTACGAGACGACTGTCCAACTCAAGCATTTGTGTATGAGGCGAGCGAGTATC 660
Qy 1002 ATGCATATCCCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1061
Db 661 ATGCATATCCCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 720

QY 1062 GCGCTCACTCCACCGCTCGCGCCAGGAAATGCCAGCGCTCCCACTACGACAAATACGAGC 1121
 Db 721 GCGCTCACTCCACCGCTCGCGCCAGGAAATGCCAGCGCTCCCACTACGACAAATACGAGC 780
 QY 1122 CACGTCGACTGCTGCTGTTGGGAGCGCTGCTTCTGCTCCGCTATGTAAGTGGGGGATCTC 1181
 Db 781 CACGTCGACTGCTGCTGTTGGGAGCGCTGCTTCTGCTCCGCTATGTAAGTGGGGGACCTC 840
 QY 1182 TCGGATCATTTTCTGCTCTCCAGCTGTTTCACTTCTGCTCTGCGCTCGCGGCATGAGACA 1241
 Db 841 TCGGATCATTTTCTGCTCTCTCAGCTGTTTCACTTCTGCTCTGCGCTCGCGGCATGAGACA 900
 QY 1242 GTGAGAGCTGCAACTGCTCAATCTATCCCGCATGATATCAGGTCACCGATGCTTGG 1301
 Db 901 TTAAGGACTGTAATGCTCAATTTATCCCGGCCATGTCGCGGTACCGTATGCTTGG 960
 QY 1302 GATATGATGAATGCTGCTCACTCAACAGCGCTAGTGGTGTGCGAGTTGCTCGGATC 1361
 Db 961 GACATGATGAATGCTGCTGCGCCCAACAACAGCGCTAGTGGTGTGCGAGTTTACTCGGATC 1020
 QY 1362 CCACAGCTGCTGCAATGCTGCGGGGGGCGCACTGGGAGTCTCTGCGGGCTTGGC 1421
 Db 1021 CCACAGCTGCTGCAATGCTGCGGGGGGCGCACTGGGAGTCTCTGCGGGCTTGGC 1080
 QY 1422 TACTATTCATGCTAGGAACTGGGCTAAGGTTCTGATTGTCGCTACTCTTTGCCGGC 1481
 Db 1081 TACTATTCATGCGGGGAACTGGGCTAAGGTTCTGATTGTCGCTACTCTTTGGCTGGC 1140
 QY 1482 GTTGACGGGAGACCAACAGAGGGAGGTTGGCGGCCCAACACCACTTCGGGTTACG 1541
 Db 1141 GTTGACGGGAGTACCACTGTCAGGGGGGGGCAAGCAAAACCAACCAAGGCTCGT 1200
 QY 1542 TCCCTTTTCTCATCTGGGGGCTCTCAGAAATCCAGCTGTGAAATACCAACGGAGCTGG 1601
 Db 1201 TCCATGTTCCGAAGTGGGCGCTCTCAGAAATCCAGCTGTGAAATACCAACGGAGTGG 1260
 QY 1602 CACATCAACAGAGCTGCTCAATTAATGCAATGCACTCCCTCCAACTCGGGTTCTTTGCCGG 1661
 Db 1261 CACATCAACAGAGCTGCTCAATGCAATGCACTCTCTCCAGCTGGGTTCTTTGCCGG 1320
 QY 1662 CTGTTTACGACACAAGTTCACTCTGTCGGGTCGCCGGAGCGATGCCAGCTGCCGC 1721
 Db 1321 CTGTTTACGACACAAGTTTCACTCTGTCGGGTCGCCGGAGCGATGCCAGCTGCCGC 1380
 QY 1722 CCATATGACTGCTCGCCAGGGTGGGCGCCCATCACTACTAAGCTAAGCTAAGCTG 1781
 Db 1381 ACCATGACAGTTTCCAGCGGATGGGTCCTCATTAATCTTATGCTGAGTCTAGCAGATCA 1440
 QY 1782 GATCAGAGCTTATGTTGGCAATTAAGCGCTCGACCGTGTGCTGTGTAACCCCGTGG 1841
 Db 1441 GACCAAGGCGCATATTGCTGGCACTACCCACCTCCACAATGTAACCATGTAACCTGGTGG 1500
 QY 1842 CAGGTGTGGTCCAGTGTATGTTTCAACCAAGCGCTGTGTTGGTGGGACCAACGAT 1901
 Db 1501 GAGGTGTGGGCGCGCTGCTGCTTCAACCAAGCGCTGTGTTGGTGGGACCAACGAT 1560
 QY 1902 CGTTCGGGTGCTTACGTATGCTGGGGGAGATGACAGACGCTGATGCTCTCAAC 1961
 Db 1561 CGTTCGGGTGCTTACGTATGCTGGGGGAGACGAGACTGACGCTGCTGCTCAAC 1620
 QY 1962 AACAGCTGCGGCACAGGCACTGGTTCGGCTGTACATGGATGAATAGTACTGGTTC 2021
 Db 1621 AACAGCTGCGGCACAGGCACTGGTTCGGCTGTACATGGATGAATAGTACTGGTTC 1680
 QY 2022 ACTACAGCTGCGGAGTTCCTCGGTGTACATCGGGGGGTGGTAAACCGACCTGTATC 2081
 Db 1681 ACCAAGACATGCGGGGGGCGCGGTGTAACTCGGGGGGTGGGACCAACACCTGTGAC 1740
 QY 2082 TGCCCCACGAGCTGCTTCCGGAAGACCCCGAGGCTACTTACACAAATGTGGCTCGGG 2141
 Db 1741 TGCCCCACGAGCTGCTTCCGGAAGACCCCGAGGCTACTTACACAAATGTGGTTCGGGG 1800

QY 2142 CCCTGGTTGACACTAGTGCCTAGTAGACTACCATACAGGCTTTGGCACTACCCCTGC 2201
 Db 1801 CCTTGGCTGACACTAGTGCCTAGTGCCTAGTGCCTAGTGCCTAGTGCCTAGTGCCTAG 1860
 QY 2202 ACTCTCAATTTTCCATCTTAAAGTTAGATGTATGCGGGGCGTGGAGCACAGGCTC 2261
 Db 1861 ACTGTAACTTACCATCTTCAAGTTAGATGTATGCGGGGGGGTGGAGCACAGGCTC 1920
 QY 2262 ANTGCCCATGCAATTTGACTCGAGGAGGCTGTAACTTGGAGGACAGGATAGTCA 2321
 Db 1921 ANTGCTCATGCAATTTGACCCGAGGAGCGTTGACITTTGAGGACAGGATAGGCGC 1980
 QY 2322 GAATCAGCGCGTCTGCTCTCTACAAAGAGTGGCAGATATGCGCTCTGCTTTCAAC 2381
 Db 1981 GAGCTCAGCGCGTCTGCTCTCTCTACAAAGAGTGGCAGTACTGCGCTCTCTCAAC 2040
 QY 2382 ACCCTACCGGCTTATCCATCTGTTTGAATCCATCTCCATCAGAACTATCTGGAGTCA 2441
 Db 2041 ACCCTACCGCTTCTGCTCCACTGCTGCTTGAATCCATCTCCATCAGAACTATCTGGAGTCA 2100
 QY 2442 TACCTGTAACGCTGTAAGGTTCAGCTTTTCTCTCTTTGCAATCAAATGGAGTACATCTG 2501
 Db 2101 TACCTGTAACGCTGTAAGGTTCAGCTTTTCTCTCTTTGCAATCAAATGGAGTATCTG 2160
 QY 2502 TTGCTTTTCTCTCTCTGCGAGACGCGGCTGTGCTCTCTCTGCTGCTGCTGCTGCTG 2561
 Db 2161 TTGCTTTTCTCTCTCTGCGAGCGACGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
 QY 2562 ATAGCCAGGCTGAGCGGCTTTAGAGAACTTGGTGTCTCTAATGCGGCTGCGTGGCC 2621
 Db 2221 ATAGCCAGGCTGAGCGGCTTTGAGAACTTGGTGTCTCTAATGCGGCTGCTGCTGCTG 2280
 QY 2622 GGAGGATGATATCT 2681
 Db 2281 GGAGGATGATATCT 2340
 QY 2682 AGGCTGCTCTCTGCGGCGGCTATGCTTTTATGCGCTATGCGGCTGCTCTCTCTCT 2741
 Db 2341 AGGCTGCTCTCTGCGGCGGCTATGCTTTTATGCGCTATGCGGCTGCTCTCTCTCT 2400
 QY 2742 CTGAGCTTACACACAGCTTACGCTTGGACCGGAGATGGCTGATCTGTCGGGGT 2801
 Db 2401 CTGAGCTTACACACAGCTTACGCTTGGACCGGAGATGGCTGATCTGTCGGGGT 2460
 QY 2802 GCGGTCTCTGTAAGTCTGCTGCTTCTGACCTTGTCAACACTACTACAAAGTCTTCTCACT 2861
 Db 2461 GCGGTCTCTGTAAGTCTGCTGCTTCTGACCTTGTCAACACTACTACAAAGTCTTCTCACT 2520
 QY 2862 AGGCTCATATGCTGTTACATCTTTTATACAGAGCGGAGCGACATGCAAGTGTGG 2921
 Db 2521 AGGCTCATATGCTGTTACATCTTTTATACAGAGCGGAGCGACATGCAAGTGTGG 2580
 QY 2922 GTCCCCCCCCCTCAAGTTTGGGAGGCGGATGCCATCTCTCTCTCTCTCTCTCTCTCTCT 2981
 Db 2581 ATCCCCCCCCCTCAAGCTTGGGAGGCGGATGCCATCTCTCTCTCTCTCTCTCTCTCTCT 2640
 QY 2982 CATCCAGAGTTAAATTTTGAATCACCACAACTCTCTCTCTCTCTCTCTCTCTCTCTCT 3041
 Db 2641 CATCCAGAGTTAAATTTTGAATCACCACAACTCTCTCTCTCTCTCTCTCTCTCTCTCT 2700
 QY 3042 GTGCTCCAGGCTGCGATTAACAGAGTGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3101
 Db 2701 GTGCTCCAGGCTGCGATTAACAGAGTGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
 QY 3102 GCATGCTATGTTAGTCCGAAAGTCTCGCGGGTCTATATGTCACAAATGCTCTCTCTCTCT 3161
 Db 2761 GCATGCTATGTTAGTCCGAAAGTCTCGCGGGTCTATATGTCACAAATGCTCTCTCTCT 2820
 QY 3162 CTGGGCGGCTGACAGGTACGTAAGTCTTATAACCATCTTACCCCACTCTCGGAGTGGGCC 3221
 Db 2821 CTGGGCGGCTGACAGGTACGTAAGTCTTATAACCATCTTACCCCACTCTCGGAGTGGGCC 2880
 QY 3222 CACCGGGGCTTACGAGACCTTTGCGGTGGCGGTAGAGCCCGTCTCTCTCTCTCTCTCT 3281

Db CGCGGGCCCTACGAGACCTTTGGGTGGTGGAGCCCGTCTCTTCCGACATGGAG 2940
 QY ACCAGGTTCATCACCTGGGAGCAGACACCGTGGTGGGACATCATCTTGGTCTA 3341
 Db ACCAAGATTCATCACCTGGGAGCAGACACCGCGGTGGGACATCATCTGGGTCTG 3000
 QY CCCGTCTCCGCCGGAAGGGGAGAGATATTTTGGACCGGTGATAGTCTCGAAGG 3401
 Db CCCGTCTCCGCCGGAAGGGGAAAAGAGATACTCTGGGCCCGGCGATAGTCTTGAAGG 3060
 QY CAAGGTTGGCACTCTTGGGCGCATACGAGCTTACTCCCAACAAACCGGGGGGTACTT 3461
 Db CGGGGTTGGCACTCTTGGGCGCATACGAGCTTACTCCCAACAAACCGGGCGCTACTT 3120
 QY GGTTCATCATCACTAGCTTACAGGCGCGGACAGAGACCTGAGTCTGAGGAGGTTCA 3521
 Db GGTTCATCATCACTAGCTTACAGGCGCGGACAGAGACCTGAGTCTGAGGAGGTTCA 3180
 QY GTGGTTTTCACCGCAACAATCTTCTTGGCGACCTGATCAACCGGGTGTGGTACT 3581
 Db GTGGTTTTCACCGCAACAATCTTCTTGGCGACCTGATCAACCGGGTGTGGTACT 3240
 QY GTTACCATGGGTGGTTCGAGACCTTACCGGTTCGAAAGGTTCGATCAACCGAAATG 3641
 Db GTTACCATGGGTGGTTCGAGACCTTACCGGTTCGAAAGGTTCGATCAACCGAAATG 3300
 QY TACACCATGTAGACCTTGGACCTTGGTGGCGACCTGATCAACCGGGCGGTCCATG 3701
 Db TACACCATGTAGACCTTGGACCTTGGTGGCGACCTGATCAACCGGGCGGTCCATG 3360
 QY ACACCATGAGTGGGAGCTTGGACCTTGGTGGCGACCTGATCAACCGGATGATCTATT 3761
 Db ACACCATGAGTGGGAGCTTGGACCTTGGTGGCGACCTGATCAACCGGATGATCTATT 3420
 QY CGGTGGCGCGGAGGCGAGCGAGGGAAGTCTTACTCTTCTCCCGAGCGCGTCTCTAC 3821
 Db CGGTGGCGCGGAGGCGAGCGAGGGAAGTCTTACTCTTCTCCCGAGCGCGTCTCTAC 3480
 QY CTGAAGGCTCTCGGGTGGTTCATTTGGTCTTGGCGACCTTGGGCGGTCTTCTG 3881
 Db CTGAAGGCTCTCGGGTGGTTCATTTGGTCTTGGCGACCTTGGGCGGTCTTCTG 3540
 QY CGGTCTGTGTGTCACCGGGGGTTCGGAAGCGGTGGACTTTGTGCCGTAGTGTCC 3941
 Db CGGTCTGTGTGTCACCGGGGGTTCGGAAGCGGTGGACTTTGTGCCGTAGTGTCC 3600
 QY ATGGAATCTACCATGGGTCTCCGTCTTTCAGACAACTCAACCCCGCGGTGTACCG 4001
 Db ATGGAATCTACCATGGGTCTCCGTCTTTCAGACAACTCAACCCCGCGGTGTACCG 3660
 QY CAGACATTTCCAAAGTGGCACATCTGCACGCTCTTCTTGGCGGACGTCGTTGGGCTTC 4061
 Db CAGTATTTTCAAGTGGCCCACTTACAGCTCCACTGGCGCGGACAGAGTACTTAAAGT 3720
 QY CGGTCTGTGTGTCACCGGGGGTTCGGAAGCGGTGGACTTTGTGCCGTAGTGTCC 4121
 Db CGGTCTGTGTGTCACCGGGGGTTCGGAAGCGGTGGACTTTGTGCCGTAGTGTCC 3780
 QY ACCTTAGGTTTGGGGCGTATATGTTCGAGGACAGGTATGACCTCAACCTCAGAACT 4181
 Db ACCTTAGGTTTGGGGCGTATATGTTCGAGGACAGGTATGACCTCAACCTCAGAACT 3840
 QY GGGGTAAAGACCATTTACACCGGGCGTCTTACTTCTTCTTGGCGAGTTCCTT 4241
 Db GGGGTAAAGACCATTTACACCGGGCGTCTTACTTCTTCTTGGCGAGTTCCTT 3900
 QY GCCGAGGTGGGTCTTGGGGCGCTTATGATCATATAATATGTGAGTGGCACTCA 4301
 Db GCCGATGGTGGGTCTTGGGGCGCTTATGATCATATAATATGTGAGTGGCACTCA 3960
 QY ACTGACTGACTTACCATTTTGGGCACTGGGCAAGTCTTGGACCAAGCGGAGTGGCTGA 4361

3361 ACTGACTGACTTACCATTTTGGGCACTGGGCAAGTCTCTGGAACGAGGAGACGGCTGGA 4020
 QY GCGGGGTCTGTCGTCGCGACACCGCTACACCTCCGGGATCGGTACCGTCCACACCCC 4421
 Db GCGGGGTCTGTCGTCGCGACACCGCTACACCTCCGGGATCGGTACCGTCCACACCCC 4080
 QY AATATCAGGAAATAGCGCTGTCCAAATGAGAGATCCCTTCTATGGAAGGCAATC 4481
 Db AACATCAGGAGGTGGCCCTGTCTAATCTGAGAGATCCCTTCTATGGAAGGCAATC 4140
 QY CCCATGAGGCCATCAAGGGGGGAGGCACTCTATTTCTGCGCATTCGAGAAATGT 4541
 Db CCCATGAGGCCATCAAGGGGGGAGGCACTCTATTTCTGTCTTCCGAGAAAGTGC 4200
 QY GACGAGCTCGCGCAAAAGCTGACAGGCGCTGAACTGAACTGAGTATATACCGGGG 4601
 Db GACGAGCTCGCGCAAAAGCTGACAGGCGCTGAACTGAACTGAGTATATACCGGGG 4260
 QY CTTGATGTGTCGTCATACCGGCTTATCGGAGACGTCTGTCGTCGGAACAGAGCTCTA 4661
 Db CTTGATGTGTCGTCATACCGGCTTATCGGAGACGTCTGTCGTCGGAACAGAGCTCTG 4320
 QY ATGAGCGGTTCACCGGCTTATTCATCTCATGATGATGATGATGATGATGATGATGAT 4721
 Db ATGAGCGGTTCACCGGCTTATTCATCTCATGATGATGATGATGATGATGATGATGAT 4380
 QY ATGAGCGGTTCACCGGCTTATTCATCTCATGATGATGATGATGATGATGATGATGAT 4781
 Db ATGAGCGGTTCACCGGCTTATTCATCTCATGATGATGATGATGATGATGATGATGAT 4440
 QY GCGGTGCGGCTCGCAACCGGAGGTAGAACTGCGAGGGGTAGGAGTGGCATCTACAG 4841
 Db GCGGTGCGGCTCGCAACCGGAGGTAGAACTGCGAGGGGTAGGAGTGGCATCTACAG 4400
 QY TTTGTGATCCAGAGACGCGCTCCGCGGTAGGACTGCGAGGGGTAGGAGGCACTTACAG 4901
 Db TTTGTGATCCAGAGACGCGCTCCGCGGTAGGACTGCGAGGGGTAGGAGGCACTTACAG 4560
 QY TATGACCGGGCTCTGCTTGGTATGAGCTCAAGCGCTGAGACCTCGTATAGTTCGG 4961
 Db TATGACCGGGCTCTGCTTGGTATGAGCTCAAGCGCTGAGACCTCGTATAGTTCGG 4620
 QY GCTTACCTTAACTACACCGGCTTCCGCTGTGCGAGACCACTGCGAGTTCCTGCGGAG 5021
 Db GCTTACCTTAACTACACCGGCTTCCGCTGTGCGAGACCACTGCGAGTTCCTGCGGAG 4680
 QY GTCTTCAAGGCTTACCATAGATGCGCACTTCTGTCGCGAGCTTAACTAGGAGGAG 5081
 Db GTCTTCAAGGCTTACCATAGATGCGCACTTCTGTCGCGAGCTTAACTAGGAGGAG 4740
 QY GACAACTTTCCCTTACCTGTCGTCATCAAGCTACAGTGTGCGGCGAGGCTCAAGCTCCA 5141
 Db GACAACTTTCCCTTACCTGTCGTCATCAAGCTACAGTGTGCGGCGAGGCTCAAGCTCCA 4800
 QY CCTCATGCTGGGACCAATGTGGAAGTGTCTCATAGCGGTGAAACCTTACCTGCAAGG 5201
 Db CCTCATGCTGGGACCAATGTGGAAGTGTCTCATAGCGGTGAAACCTTACCTGCAAGG 4860
 QY CCACACCGCTTGTATAGCTTGGAGCGCTCCGAGGAGGCTGAGGAGGAGGAGGAGGAG 5261
 Db CCACACCGCTTGTATAGCTTGGAGCGCTCCGAGGAGGCTGAGGAGGAGGAGGAGGAG 4920
 QY ATAACTAAATACATCATGCGATGATGTCGCTGACCTGAGGAGGCTGCTACTAGACCTGG 5321
 Db ATAACTAAATACATCATGCGATGATGTCGCTGACCTGAGGAGGCTGCTACTAGACCTGG 4980
 QY GTGCTGGTAGGCGAGTCTTGGAGCTTGGCGCGCTTGGCGCGCTTGGCGCGCTTGGCG 5381
 Db GTGCTGGTAGGCGAGTCTTGGAGCTTGGCGCGCTTGGCGCGCTTGGCGCGCTTGGCG 5040
 QY GTCAATGTGGCGAGGATCATCTTGTCCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5441
 Db GTCAATGTGGCGAGGATCATCTTGTCCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5100

